GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2002, 20:31:42 ; Search time 1818.72 Seconds
(without alignments)
736.085 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 227
Sequence: 1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105348_4834/app_query.fasta_1.398
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017_@CGN_1_1_11616_@runat_05112002_105348_4834 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
```
1:  gb_ba:*        29: em_vi:*
2:  gb_htg:*       30: em_htg_hum:*
3:  gb_in:*        31: em_htg_inv:*
4:  gb_om:*        32: em_htg_other:*
5:  gb_ov:*        33: em_htg_mus:*
6:  gb_pat:*       34: em_htg_pln:*
7:  gb_ph:*        35: em_htg_rod:*
8:  gb_pl:*        36: em_htg_mam:*
9:  gb_pr:*        37: em_htg_vrt:*
10: gb_ro:*        38: em_sy:*
11: gb_sts:*       39: em_htgo_hum:*
12: gb_sy:*        40: em_htgo_mus:*
13: gb_un:*        41: em_htgo_other:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	
	1	227	100.0	141	1	AF277152	AF277152	Streptoco
	2	227	100.0	141	1	AF277153	AF277153	Streptoco
	3	227	100.0	141	1	AF277155	AF277155	Streptoco
	4	227	100.0	141	1	AF277156	AF277156	Streptoco
	5	227	100.0	141	1	AF277157	AF277157	Streptoco
	6	220	96.9	141	1	AF277151	AF277151	Streptoco
	7	211	93.0	159	1	AF277154	AF277154	Streptoco
c	8	69	30.4	8411	6	BD003696	BD003696	Polynucle
	9	69	30.4	10477	1	AE007510	AE007510	Streptoco
	10	69	30.4	10478	1	AE008567	AE008567	Streptoco
c	11	69	30.4	232807	2	SPNEU1901	AL449923	Streptoco
c	12	67	29.5	116557	2	AC107022	AC107022	Homo sapi
c	13	67	29.5	149335	2	AC009651	AC009651	Homo sapi
c	14	67	29.5	174028	2	AC067831	AC067831	Homo sapi
c	15	67	29.5	180389	2	AC063944	AC063944	Homo sapi
	16	65.5	28.9	173868	2	AP001525	AP001525	Homo sapi
	17	65.5	28.9	181083	2	AP001268	AP001268	Homo sapi
c	18	64.5	28.4	301550	1	AP003192	AP003192	Clostridi
c	19	64	28.2	118226	9	HS46618	AL030998	Homo sapi
	20	63.5	28.0	2455	1	LLJ002203	AJ002203	Lactococc
	21	63.5	28.0	10899	1	AF242367	AF242367	Lactococc
c	22	63.5	28.0	13028	3	AC024836	AC024836	Caenorhab
c	23	63.5	28.0	41385	3	AF106573	AF106573	Caenorhab
c	24	63.5	28.0	87246	9	AC002539	AC002539	Homo sapi
	25	63.5	28.0	133925	9	AC004562	AC004562	Homo sapi
	26	63.5	28.0	136421	2	AC006883	AC006883	Caenorhab
	27	63.5	28.0	150789	2	AC006885	AC006885	Caenorhab
c	28	63.5	27.5	155609	9	HSJ813016	AL049823	Human DNA
	29	62.5	27.5	87863	9	AL513188	AL513188	Human DNA
	30	62.5	27.5	153762	9	AL355338	AL355338	Human DNA
	31	62.5	27.5	171031	2	AC104020	AC104020	Homo sapi
c	32	62.5	27.5	185967	9	CNS01DVR	AL136001	Human chr
	33	62.5	27.3	213087	2	AC124973	AC124973	Mus muscu
	34	62	27.3	110000	2	CEY105E8_2	Continuation (3 of	
c	35	62	27.3	128746	2	AC094590	AC094590	Rattus no
	36	62	27.3	162427	2	AC115502	AC115502	Rattus no
	37	62	27.3	187260	9	AC018976	AC018976	Homo sapi
	38	62	27.3	277607	3	CEY105E8A	AL132876	Caenorhab
	39	61.5	27.1	157381	9	AL360267	AL360267	Human DNA
	40	61.5	27.1	176085	9	AL391600	AL391600	Human DNA
	41	61.5	27.1	298406	3	CEY75B8A	AL033514	Caenorhab
	42	61	26.9	135382	9	AC023798	AC023798	Homo sapi
	43	61	26.9	187732	9	CNS01RHM	AL161851	Human chr
c	44	61	26.9	229547	2	AC063935	AC063935	Homo sapi
	45	61	26.9	283274	2	AC055759	AC055759	Homo sapi

ALIGNMENTS

RESULT 1

```
LOCUS       AF277152               141 bp    DNA     linear   BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain GB14 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277152
VERSION     AF277152.1  GI:12698429
KEYWORDS    .
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
  PUBMED    11208787
REFERENCE   2 (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="GB14"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01542.1"
                     /db_xref="GI:12698430"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a   19 c   25 g   39 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                      1.01e-25        Length:         141
Percent Similarity:         227.00          Matches:        46
Best Local Similarity:      100.00%         Conservative:   0
DB:                         100.00%         Mismatches:     0
                            100.00%         Indels:         0
                            1               Gaps:           0

US-09-833-017B-2 (1-46) x AF277152 (1-141)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy   41 ThrGlnAlaLeuGlyLys 46
        ||||||||||||||||||
Db  121 ACACAAGCTTTGGGAAAA 138

RESULT 2
LOCUS       AF277153
DEFINITION  Streptococcus mutans strain H7 competence stimulating protein
            precursor (comC) gene, complete cds.
```

```
ACCESSION   AF277153
VERSION     AF277153.1  GI:12698431
KEYWORDS    .
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
  PUBMED    11208787
REFERENCE   2 (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="H7"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01543.1"
                     /db_xref="GI:12698432"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a   19 c   25 g   39 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                      1.01e-25        Length:         141
Percent Similarity:         227.00          Matches:        46
Best Local Similarity:      100.00%         Conservative:   0
DB:                         100.00%         Mismatches:     0
                            100.00%         Indels:         0
                            1               Gaps:           0

US-09-833-017B-2 (1-46) x AF277153 (1-141)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy   41 ThrGlnAlaLeuGlyLys 46
        ||||||||||||||||||
Db  121 ACACAAGCTTTGGGAAAA 138

RESULT 3
LOCUS       AF277155
ACCESSION   AF277155
DEFINITION  Streptococcus mutans strain LT11 competence stimulating protein
            precursor (comC) gene, complete cds.
VERSION     AF277155.1  GI:12698435
KEYWORDS    
SOURCE      Streptococcus mutans.
```

```
ORGANISM  Streptococcus mutans
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE 1 (bases 1 to 141)
AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE     Natural genetic transformation of Streptococcus mutans growing in
          biofilms
JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
MEDLINE   21142515
PUBMED    11208787
REFERENCE 2 (bases 1 to 141)
AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
TITLE     Direct Submission
JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
          124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="LT11"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01545.1"
                     /db_xref="GI:12698436"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a    19 c    25 g    39 t
ORIGIN

Alignment Scores:
Pred. No.:              1.01e-25          Length:            141
Score:                  227.00            Matches:           46
Percent Similarity:     100.00%           Conservative:      0
Best Local Similarity:  100.00%           Mismatches:        0
Query Match:            100.00%           Indels:            0
DB:                     1                 Gaps:              0

US-09-833-017B-2 (1-46) x AF277155 (1-141)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACATCATTCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy   41 ThrGlnAlaLeuGlyLys 46
        ||||||||||||||||||
Db  121 ACACAAGCTTTGGGAAAA 138

RESULT 4
AF277156
LOCUS       AF277156               141 bp    DNA     linear   BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain NG8 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277156
VERSION     AF277156.1  GI:12698437
KEYWORDS    .
SOURCE      Streptococcus mutans
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
```

```
AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE     Natural genetic transformation of Streptococcus mutans growing in
          biofilms
JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
MEDLINE   21142515
PUBMED    11208787
REFERENCE 2 (bases 1 to 141)
AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
TITLE     Direct Submission
JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
          124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="NG8"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01546.1"
                     /db_xref="GI:12698438"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a    19 c    25 g    39 t
ORIGIN

Alignment Scores:
Pred. No.:              1.01e-25          Length:            141
Score:                  227.00            Matches:           46
Percent Similarity:     100.00%           Conservative:      0
Best Local Similarity:  100.00%           Mismatches:        0
Query Match:            100.00%           Indels:            0
DB:                     1                 Gaps:              0

US-09-833-017B-2 (1-46) x AF277156 (1-141)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheGluIleLysThrAspGluLeuGlu 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACATCATTCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy   41 ThrGlnAlaLeuGlyLys 46
        ||||||||||||||||||
Db  121 ACACAAGCTTTGGGAAAA 138

RESULT 5
AF277157
LOCUS       AF277157               141 bp    DNA     linear   BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain UA159 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277157
VERSION     AF277157.1  GI:12698439
KEYWORDS    
SOURCE      Streptococcus mutans
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE     Natural genetic transformation of Streptococcus mutans growing in
          biofilms
JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
```

```
                   TITLE        Direct Submission
                   JOURNAL      Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
                                124 Edward Street, Toronto, ON M5G 1G6, Canada
                   FEATURES                  Location/Qualifiers
                         source           1..141
                                          /organism="Streptococcus mutans"
                                          /strain="BM71"
                                          /db_xref="taxon:1309"
                         gene             1..141
                                          /gene="comC"
                         CDS              1..141
                                          /gene="comC"
                                          /note="ComC"
                                          /codon_start=1
                                          /transl_table=11
                                          /product="competence stimulating protein precursor"
                                          /protein_id="AAK01541.1"
                                          /db_xref="GI:12698428"
                                          /translation="MKKTPSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                                          GK"
                         mat_peptide      76..138
                                          /gene="comC"
                                          /product="competence stimulating protein"
                   BASE COUNT       58 a       20 c      25 g      38 t
                   ORIGIN

                   Alignment Scores:
                   Pred. No.:              1.14e-24         Length:            141
                   Score:                  220.00           Matches:            45
                   Percent Similarity:     97.83%           Conservative:        0
                   Best Local Similarity:  97.83%           Mismatches:          1
                   Query Match:            96.92%           Indels:              0
                   DB:                     1                Gaps:                0

                   US-09-833-017B-2 (1-46) x AF277151 (1-141)

                   Qy     1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
                             |||||||||||                                                          ||||||
                   Db     1 ATGAAAAAACACCATCATTAAAAATGACTTTAAGAAATTAAGACTGATGAATTAGAG 60

                   Qy    21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
                             |||||||||||||||||||||||||||||||||||||||||
                   Db    61 ATTATCATTGGCGGAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

                   Qy    41 ThrGlnAlaLeuGlyLys 46
                             |||||||||||||||
                   Db   121 ACACAAGCTTTGGGAAAA 138

                   RESULT 7
                   AF277154
                   LOCUS       AF277154               159 bp    DNA    linear   BCT 07-FEB-2001
                   DEFINITION  Streptococcus mutans strain JH1005 competence stimulating protein
                               precursor (comC) gene, complete cds.
                   ACCESSION   AF277154
                   VERSION     AF277154.1  GI:12698433
                   KEYWORDS    .
                   SOURCE      Streptococcus mutans.
                     ORGANISM  Streptococcus mutans
                               Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
                               Streptococcus.
                   REFERENCE   1  (bases 1 to 159)
                     AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
                     TITLE     Natural genetic transformation of Streptococcus mutans growing in
                               biofilms
                     JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
                     MEDLINE   21142515
                     PUBMED    11208787
                   REFERENCE   2  (bases 1 to 159)
                     AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
                     TITLE     Direct Submission
                     JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
                               124 Edward Street, Toronto, ON M5G 1G6, Canada
                   FEATURES                  Location/Qualifiers
```

```
MEDLINE     21142515
PUBMED      11208787
REFERENCE   2  (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES                  Location/Qualifiers
      source           1..141
                       /organism="Streptococcus mutans"
                       /strain="UA159"
                       /db_xref="taxon:1309"
      gene             1..141
                       /gene="comC"
      CDS              1..141
                       /gene="comC"
                       /note="ComC"
                       /codon_start=1
                       /transl_table=11
                       /product="competence stimulating protein precursor"
                       /protein_id="AAK01547.1"
                       /db_xref="GI:12698440"
                       /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                       GK"
      mat_peptide      76..138
                       /gene="comC"
                       /product="competence stimulating protein"
BASE COUNT       58 a       19 c      25 g      39 t
ORIGIN

Alignment Scores:
Pred. No.:              1.01e-25         Length:            141
Score:                  227.00           Matches:            46
Percent Similarity:     100.00%          Conservative:        0
Best Local Similarity:  100.00%          Mismatches:          0
Query Match:            100.00%          Indels:              0
DB:                     1                Gaps:                0

US-09-833-017B-2 (1-46) x AF277157 (1-141)

Qy     1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGAAAAAACACTATCATTAAAAATGACTTTAAGAAATTAAGACTGATGAATTAGAG 60

Qy    21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ATTATCATTGGCGGAGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy    41 ThrGlnAlaLeuGlyLys 46
          ||||||||||||||||||
Db   121 ACACAAGCTTTGGGAAAA 138

RESULT 6
AF277151
LOCUS       AF277151               141 bp    DNA    linear   BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain BM71 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277151
VERSION     AF277151.1  GI:12698427
KEYWORDS    .
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
  PUBMED    11208787
REFERENCE   2  (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
```

```
     source          1..159
                     /organism="Streptococcus mutans"
                     /strain="JH1005"
                     /db_xref="taxon:1309"
     gene            1..132
                     /gene="comC"
     CDS             1..132
                     /gene="comC"
                     /note="comC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01544.1"
                     /db_xref="GI:12698434"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGTLSTFFRLFNRSFTQA"
     mat_peptide     76..129
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT     65 a     23 c     26 g     45 t
ORIGIN

Alignment Scores:
Pred. No.:              3e-23
Score:                  211.00
Percent Similarity:     88.46%
Best Local Similarity:  86.54%
Query Match:            92.95%
DB:                     1

US-09-833-017B-2 (1-46) x AF277154 (1-159)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACTTAGTCTCATTAAAAATGACTTTAAAGAAATTAAAACTGATGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerThrPhePheArgLeuPheAsnArgSer-- 39
        ||||||||||||||||||||||||||||||||::|||||||||||||||||||
Db   61 ATTATCATTGGCGGAAGCGGAACCCGTCAACATTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy   40 ----------PheThrGlnAlaLeuGlyLys 46
                   ||||||||||||||||||
Db  121 ACACAAGCTAGAATTTTACACAGCTTTGGGAAAA 156

RESULT 8
BD003696/c
LOCUS       BD003696        8411 bp    DNA    linear    PAT 31-JAN-2002
DEFINITION  Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION   BD003696
VERSION     BD003696.1  GI:18631657
KEYWORDS    JP 2001501833-A/16.
SOURCE      unidentified
ORGANISM    unidentified
            unclassified.
REFERENCE   1  (bases 1 to 8411)
AUTHORS     Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
            Fannon,M. and Dougherty,B.A.
TITLE       Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL     Patent: JP 2001501833-A 16 13-FEB-2001;
            HUMAN GENOME SCIENCES INC
COMMENT     OS   Unidentified
            PN   JP 2001501833-A/16
            PD   13-FEB-2001
            PF   30-OCT-1997 JP 1998520718
            PR   31-OCT-1996 US 60/029960
            PI   CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
                 STEVEN C BARASH,
            PI   MICHAEL FANNON,BRIAN A DOUGHERTY
            PC   C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
            PC   C12N1/21,
            PC   C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N5/00,C12N5/00, PC
            G06F15/40
            CC   Strandedness: Double;
            CC   Topology: Linear;
```

```
FH  Key             Location/Qualifiers
FT  source          1..8411
FT                  /organism='Unidentified'.

FEATURES             Location/Qualifiers
     source          1..8411
                     /organism="unidentified"
                     /db_xref="taxon:32644".

BASE COUNT     2741 a   1386 c   1639 g   2645 t
ORIGIN

Alignment Scores:
Pred. No.:              8.68
Score:                  69.00
Percent Similarity:     50.88%
Best Local Similarity:  33.33%
Query Match:            30.40%
DB:                     6

US-09-833-017B-2 (1-46) x BD003696 (1-8411)

Qy    2 LysLysThrLeuSerLeuLysAsnAspPhe---------------------LysGlu 13
        ||||||||:::                                         |||||
Db 3566 AAAAAGACTTTACAACTGATAATGCTTCTCCCTCTTTAATCTAGGATATACTAAAGAG 3507

Qy   14 IleLysThrAspGluLeuGluIleIleIleGlyGlySerGlySerLeuSer--------- 30
        ||||||                  :::      :::            ::|||
Db 3506 AAAAAAACTTTCAATATCAATGTTAGTTTCCTGGAAATTCACAAGTACATTTGAATCT 3447

Qy   31 ---ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 46
           :::|||:::|||||          ::::::|||||||:::::||
Db 3446 CCTACCTTCTCATCGTTTAGATACCAAAACTTTCACCGAGCGCAATTCAAAAA 3396

RESULT 9
AE007510
LOCUS       AE007510        10477 bp    DNA    linear    BCT 31-AUG-2001
DEFINITION  Streptococcus pneumoniae TIGR4    section 193 of 194 of the complete
            genome.
ACCESSION   AE007510 AE005672
VERSION     AE007510.1  GI:14973736
KEYWORDS    .
SOURCE      Streptococcus pneumoniae TIGR4.
ORGANISM    Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10477)
AUTHORS     Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
            Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,
            McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.
TITLE       Complete genome sequence of a virulent isolate of Streptococcus
            pneumoniae
JOURNAL     Science 293 (5529), 498-506 (2001)
MEDLINE     21357209
PUBMED      11463916
REFERENCE   2  (bases 1 to 10477)
AUTHORS     Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
            Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
            Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
            Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
            Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..10477
```

```
gene
        /organism="Streptococcus pneumoniae TIGR4"
        /strain="TIGR4"
        /db_xref="taxon:170187"
        /gene="SP2226"
        110..478
CDS
        110..478
        /gene="SP2226"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein"
        /protein_id="AAK76274.1"
        /db_xref="GI:14973737"
        /translation="MEYKLFEEFITLQALFKELGITHSGGAIKSFLSEHSVYFNRELE
        SRRGKKLRIGDKVDIPDMNIDILLTQPTSEEQDEYQADKVEKERIAKLVKKMNKGVKK
        DKSKPASSPKSKQQARPFPGR"

gene
        481..1578
        /gene="SP2227"
CDS
        481..1578
        /gene="SP2227"
        /codon_start=1
        /transl_table=11
        /product="recF protein"
        /protein_id="AAK76275.1"
        /db_xref="GI:14973738"
        /translation="MWLQHLSLKTFRNYKE?KIDFNPKLNVFLGRNAQGKTNMLEAIY
        FLALTRSHRTRTDKNLIHFDEEQLHLSGLVQKKTGSIPLEITLOKGRVTKVNHLKQA
        RLSDYVGHMNVLFAPEDLQLIKGAPSIRRKFIDMELGQIKPIYLSDLTNYNHILKQR
        NTYLKSAQKIDEFLSVLDDQLVDYGCRVMNRLDFIKKLESFGRKKHFELSNQIEEL
        SISYQSSVNITDKQNLSBSFKIALEKSRSRDLFKKNTGVGPHRDDISFYINGMDASFG
        SQQHRSLVLSIKLAEIELMESITTESPILLDDVMSELDNTRQLKLLETISQSIQTF
        ITTTSLDHLQNLPENLSIFTIQDGKAAVNGN"

gene
        complement(1629..3107)
        /gene="SP2228"
CDS
        complement(1629..3107)
        /gene="SP2228"
        /note="identified by match to PFAM protein family HMM
        PF00478"
        /codon_start=1
        /transl_table=11
        /product="inosine-5'-monophosphate dehydrogenase"
        /protein_id="AAK76276.1"
        /db_xref="GI:14973739"
        /translation="MSNWDTKFLKKGFTFDDVLLIPAESHVLPNDADLTKLADNLTL
        NIPIITAAMDTVTESQMAIAIARAGGLGVIHKNMSIAQQADEVRKVKRSENGVIIDPF
        FLIPEHTIAEADELMGRYRISGVPVVETLENRKLVGILTNRDLRFISDYNQPISNHMT
        SENLVTAPVGTDLATAEESILQEBHRIEKLPLVDEEGSLSGLITIKDIEKVIEFPNAAKD
        EFGRLLVAGAVGVTSDTFERAEALFEAGADAIVIDTAHGSAGVLRKIAEIRAHFPDR
        TLIAGNIATAEGARALYEAGVDVVKVGIGPGSICTRVIAGVGVPQVTAIYDAAAVAR
        EYGKTIIADGGIKYSGDIVKALAAGGNAVMLGSMFAGTDEAPGETEIFQGKFKTYRG
        MGSIAAMKGSSDRYFOGSVNEANKLVPEGIEGRVAYKGAAADIVFQMICGIRSGMGY
        CGAANLKELHDNAQFIEMSGAGLKESHPHDVQITNEAPNYSM"

gene
        complement(3259..4284)
        /gene="SP2229"
CDS
        complement(3259..4284)
        /gene="SP2229"
        /note="identified by match to TIGR protein family HMM
        TIGR00233"
        /codon_start=1
        /transl_table=11
        /product="tryptophanyl-tRNA synthetase"
        /protein_id="AAK76277.1"
        /db_xref="GI:14973740"
        /translation="MTKFIILTGDRPTGKLHIGHYVGSLKNRVLLQEEDKYDMFVFLA
        DQQALTDHAKDPQTIVESIGNVALDYLAVGLDPNKSTIFIQSQIPELAELSMYYMNLV
        SLARLERNPTVKTEISOKGFGESIPTGFLVPIAQAADITAFKANYVPVGTDQKPMIE
        QTREIVRSSTNNAYNCDVLVEPEGIYPENERAGRLPGLDGNAKMSKSLNGIYLADDAD
        TLRKKVMSMYTDPDHIRVEDPGKIEGNMVFHYLDVFGRPEDDAQEIADMKERYQRGGLG
        DVKTKRYLEILERELGPIRERRIEFAKDMGEVYNMIQKGSERAREVAGQTLSEVKGA
        MGLHFN"

gene
        4490..6112
        /gene="SP2230"
CDS
        4490..6112
        /gene="SP2230"
        /codon_start=1
        /transl_table=11
        /product="ABC transporter, ATP-binding protein"
        /protein_id="AAK76278.1"
        /db_xref="GI:14973741"
        /translation="MLIVSDVSLRFSDRKLFDDVNIKFTEGNTYGLIGANGAGKSTFL
        KILAGDIEPTTGHISLGPDERLSVLRQNHFDYEDERAIDVVIMGNEKLYSIMKEKDAI
        YMKEDFSDEDGVRAAELEGEFAELGGWEAESEASQLLQNLNIPEELHYQNMSELANGE
        KVKVLLAKALFGKPDVLLLDEPTNGLDIQSITWLEDFLIDFDNTVIVVSHDRHFLNKV
        CTHMADLDFGKIKLYVGNYDFWKESSELAAKLLADRNAKAEEKIKQLQEFVARFSANA
        SKSRQATSRKKMLDKIELEEIVPSSRKYPFINFKAERRIGNDLLTVENLIVKIDGETI
        LDNISFILRPDDKTALIGQNDIQTTALIRAIMGDIDYEGTVKWGVTTSQSYLPKDNSA
        DFAGGESILDWLRQFASKEEDDNTFLRGFLGRMLFSGDEVNKPVNVLSGEKVRVMLS
        KLMLLKSNVLVLDDPTNHIDLESISSLNDGLKNFKESIIFASHDHEFIQTLANHIIVL
        SKNGVIDRIDETYDEFLENAEVQAKVELWKD"

gene
        6174..8726
        /gene="SP2231"
CDS
        6174..8726
        /gene="SP2231"
        /codon_start=1
        /transl_table=11
        /product="ABC transporter, permease protein, putative"
        /protein_id="AAK76279.1"
        /db_xref="GI:14973742"
        /translation="MKSFLKTYRTYFISFIIPVVIMSGVYLSQSIYWNSDNSPLLGDG
        FHQYVIFDVALRNLRNGMHFTFTSGLSNLFKFQSIPQILKLALSTSYALMSFTVSQLEIKTWLD
        DAIYLTLLKFGIGLSTFFSLNKLFQSIPQILKLALSTSYALMSFTVSQLEIKTWLD
        VFILIPLIITGLHLLITEKKLLLYFTSLSILFIQNYFGYMTVLFLIFWYLCQISWDF
        KTRKSSVLDFIVISFLAGMASLIMTLPFLDLQTHGEKLTEVTKFQTESSWYLDLFAK
        QFIGSFDTTKYGAIPMIFVGLPPFILTILFFTLKSIKFHVKLIYVIFFAFLIASFYIE
        ALDLFWQGMHTPNMFLHRYAWIFSTLLIYTAAEVLRKLKELKVWNFLVSLFLVVAGFL
        ATIYLKSHYSFLTDLNILTLEFLVVYSLLLAVIKFISVNLFAILISLFLVEMSL
        NASSQMDGIAKEWGFASRSAYSRDIPAMESFSTYIGNQFTRTEKLQTOTGNDSMKFNY
        NGISQFSSVRNRSSSSTLDKLGFKSSGTNLNLRYANNSILADSLFGIQYNLSDSPIDK
        YGFKDIYQKDNLTIYENQYSLFIAVASQSVYNVKFNEHTLGNQASFLNQLANVNFDY
        FSPIPYEKTEKIENTNDLISVTSSNEDAAIQYQIEVPENSQVYLSFINLHFSNDKQK
        KVDILVNGEKKTFTTDNVFSFFNLGYTKEKKTFNINVSFPGNSQVSFESPTFYRLDTK
        TFTEAIQKIKEQPVTVSTSKNKVFATVDVQQDTSIFFTIPYDKGWSAYQDGKKIEIKQ
        AQTGFMKVDIPKGKGTTLSFIPNGFITGAICSFTSLLFGIYNHRRKSSKA"

gene
        complement(8827..9506)
        /gene="SP2232"
        /note="This region contains an authentic frame shift and
        is not the result of a sequencing artifact; identified by
        Glimmer2; putative; conserved hypothetical protein,
        authentic frameshift"

gene
        complement(9591..10358)
        /gene="SP2233"
CDS
        complement(9591..10358)
        /gene="SP2233"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK76280.1"
        /db_xref="GI:14973743"
        /translation="MIGISLIPDLYNIIFLSSMWDPYGQLSDLPVAVVNNDKEASYNG
        NTMAIGKDMVSNLKENKDLDFHFVDEEEGKKGLEDGDYYMVVTLPSDLSEKTTLSNI
        QSTAAYQSLITSEQQTEISDSVSONSTDSIQSAQSIVALVQDLQGSLENLQNQSSNLST
        LKNQSNQVSPITSTSLIGLSSGLTEIQGDVTSKLVPASQSIASGVNAYTGVDKVSQG
        ASQLSEKNATLTGSLDKLVSGSNTLTQKSSRLTAGVG"

BASE COUNT    3366 a   2047 c   1699 g   3365 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  11.3          Length:           10477
Percent Similarity:     69.00%        Matches:          19
Best Local Similarity:  50.88%        Conservative:     10
Query Match:            33.33%        Mismatches:       16
DB:                     30.40%        Indels:           12
                        1             Gaps:             2

US-09-833-017B-2 (1-46) x AE007510 (1-10477)
```

```
Qy        2 LysLysThrLeuSerLeuLysAsnAspPhe------------------LysGlu 13
            ||||||||     |||  |||                            |||||
Db     8247 AAAAAGACTTTTACAACTGATAATGTCTTCCTTCCTTCTTAACTCTAGGATATACTAAAGAG 8306

Qy       14 IleLysThrAspGluLeuGluIleIleGlyGlySerGlySerLeuSer-------- 30
            |||||             ||:   ::   |:  :|||               :::|||
Db     8307 AAAAAACCTTTCAATATCAATGTTAGTTTCCCTGGAAATTCACAAGTATCATTTGAATCT 8366

Qy       31 ---ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 46
            :::::::|||||||||  ::::::||||||||:::|:::  |||
Db     8367 CCTACCTTCTCATCGTTTAGATACCAAAAACTTTCACCGAGGCAATTCAAAAA 8417

RESULT 10
AE008567
LOCUS          AE008567              10478 bp    DNA      linear    BCT 13-SEP-2001
DEFINITION     Streptococcus pneumoniae R6 section 183 of 184 of the complete
               genome.
ACCESSION      AE008567 AE007317
VERSION        AE008567.1  GI:15459737
KEYWORDS       .
SOURCE         Streptococcus pneumoniae R6.
ORGANISM       Streptococcus pneumoniae R6
               Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
               Streptococcus.
REFERENCE   1 (bases 1 to 10478)
AUTHORS        Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
               DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
               Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
               LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
               McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
               Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
               Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
               Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
               and Glass,J.I.
TITLE          Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL        J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE        21429245
PUBMED         11544234
REFERENCE   2 (bases 1 to 10478)
AUTHORS        Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
               DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
               Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
               LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
               McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
               Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
               Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
               Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
               and Glass,J.I.
TITLE          Direct Submission
JOURNAL        Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
               Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES             Location/Qualifiers
     source         1..10478
                    /organism="Streptococcus pneumoniae R6"
                    /strain="R6"
                    /db_xref="taxon:171101"
     gene           110..478
                    /gene="spr2031"
     CDS            110..478
                    /gene="spr2031"
                    /codon_start=1
                    /transl_table=11
                    /label=spr2031
                    /product="Conserved hypothetical protein"
                    /protein_id="AAL00833.1"
                    /db_xref="GI:15459738"
                    /translation="MEYKLFEEFITLQALLKELGITHSGGAIKSFLSEHSVYFNRELE
                    SRRGKKLRIGDKVDIPDMNIDILLTQPTSEEQDEYGADKVEKERIAKLVKKMNKGVKK
                    DKSKPASSPKSQQAPRFPGR"
     gene           481..1578
                    /gene="recF"
                    /note="spr2032"
```



```
     CDS            481..1578
                    /gene="recF"
                    /codon_start=1
                    /transl_table=11
                    /label=spr2032
                    /product="Recombination protein RecF"
                    /protein_id="AAL00834.1"
                    /db_xref="GI:15459739"
                    /translation="MWLQHLSLKTFRNVKETKIDFNPKLNVFLGRNAQGKTNMLEAIY
                    FLALTRSHRTRPDKNLIHFDEQLHLSGLVOKTGSIPLEIELTQKGRVTKVNHLKQA
                    RLSDYVGHMNVLFAPEDLQLIKGAPSIRRKFIDMELGQIKPIYLSDLTNYNHILKQR
                    NTYLKSAQKIDETFLSVLDDQLVDYGCRVMNHRLDFIKKLESFGRKKHFELSNQIEEL
                    SISYQSSVNITDKQNLSESFKIALEKSRSRDLFKKNTGVGPHRDDISFYINGMDASFG
                    SQGQHRSLVLSIKLAEIELMESITTESPILLLDDVMSELDNTRQLKLLETISQSIQTF
                    ITTTSLDHLQNLPENLSIFTIQDGKASVNGN"
     gene           complement(1629..3107)
                    /gene="imdH"
                    /note="spr2033"
     CDS            complement(1629..3107)
                    /gene="imdH"
                    /EC_number="1.1.1.205"
                    /codon_start=1
                    /transl_table=11
                    /label=spr2033
                    /product="Inosine-5'-monophosphate dehydrogenase"
                    /protein_id="AAL00835.1"
                    /db_xref="GI:15459740"
                    /translation="MSNWDTKFLKKGFTFDDVLLIPAESHVLPNDADLTFKLADNLTL
                    NIPITAAMDTVTESQMAIAIARAGGLGVIHKNMSIAQQADEVRKVKRSENGVIIDPF
                    FLTPEHTIAEADELMGRYRISGVPVVETLENRKLVGLITNRDLRFISDYNQPISNHMT
                    SENLVTAPVGTDLATAESILQEHRIERKLPLVDEEGSLSGLVTTIKDIEKVIEPNAAKD
                    EFGRLLVAGAVGVTSDTFERAEALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPDR
                    TLIAGNIATAEGARALYEAGVDVVKVGIGPGSICTTRVIAGVGVPQVTAIYDAAVAR
                    EYGKTIIADGGIKYSGDIVKALAAGGNAVMLGSMFAGTDEAPGETEIFQGRKFTYRG
                    MGSIAAMKKGSSDRYFQGSVNRANKLVPEGIRGRVAYKGAAADIVFQMIGGIRSGMGY
                    CGAANLKELHDNAQFIEMSGAGLKESHPHDVQITNEAPNYSM"
     gene           complement(3259..4284)
                    /gene="trpS"
                    /note="spr2034"
     CDS            complement(3259..4284)
                    /gene="trpS"
                    /EC_number="6.1.1.2"
                    /codon_start=1
                    /transl_table=11
                    /label=spr2034
                    /product="Tryptophanyl-tRNA synthetase"
                    /protein_id="AAL00836.1"
                    /db_xref="GI:15459741"
                    /translation="MTKPIILTGDRPTGKLHIGHYVGSLKNRVLLQEEDKYDMFVFLA
                    DQQALTDHAKDPQTIVESIGNVALDYLAVGLDPNKSTIFIQSQIPELAELSMYYMNLV
                    SLARLERNPTVKTEISQKGFGESIPTGFLVYPIAQADITAFKANVPVGTDQKPMEIE
                    QTREIVRSFNNAYNCDVLVEPEGIYPENERAGRLPGLDGNAKMSKSLNNGIYLADDAD
                    TLRKKVMSMYTDPDHIRVEDPGKIEGNMVFHYLDVFGRPEDAQEIADMKERYQRGLG
                    DVKTKRYLLEILERELGPIRERRIEFAKDMGEVYNMIQKGSRAREVAGQTLSEVKGA
                    MGLHYFN"
     gene           4490..6112
                    /gene="ABC-NBD"
                    /note="spr2035"
     CDS            4490..6112
                    /gene="ABC-NBD"
                    /note="Putative"
                    /codon_start=1
                    /transl_table=11
                    /label=spr2035
                    /product="ABC transporter ATP-binding protein - unknown
                    substrate"
                    /protein_id="AAL00837.1"
                    /db_xref="GI:15459742"
                    /translation="MLIYVSDVSLRFSDRKLFDDVNIKFTEGNTYGLIGANGAGKSTFL
                    KILAGDIEPTTGHISLAGPRLGEERLSVLRQNHFYEDERAIDVVIMGNKLYSIMKEKDAI
                    YMKEDFSDEDGVRAAELEGEFAELGGWEAESPASQLLQNLNIPEELHYQNMSELANGE
                    KVKVLLAKALFGKPDVLLLDEPTNGLDIQSTWLEDFLIDFDNTVIVSHDRHFLNKV
                    CTHMADDLDFGKIKLYGNYDFWKESSELAAKLLADRNAKAEEKIKQLQEFVARFSANA
```

```
gene        6174. .8726
            /gene="spr2036"
CDS         6174. .8726
            /gene="spr2036"
            /codon_start=1
            /transl_table=11
            /label=spr2036
            /product="Conserved hypothetical protein"
            /protein_id="AAL00838.1"
            /db_xref="GI:15459743"
            /translation="MKSFLKIYRTYFISLLIPVVIMSGVYLSQGIYWNSDNSPLLGDG
            FHQYVIFDVALRNLHGNSSLFYTFTSGLGNFYALSSYYLGSFLAPLYFFDLTNMP
            DAVYLTTLLKFGLIGLSTFFFSLNKLFQSIPQTLKLALSTSYALMSFTVSQLEIKTWLD
            VFLIPLIITGLHLLITEKKLLLYFTSLSLSIFIQNYYFGYMTVLFLIFWYLCQISWDF
            KTRKSSVLDFIVISFLAGMASLIMTLPTLFDLQTHGEKLTEVTKFQTESSWYLDLFAK
            QFIGSFDTTKYGAIPMIFVGLFPFILTILFFTLKSIKFHVKLIYVIFFAFLIASFYIE
            ALDLFWQGMHTPNMFLHRYAWIFSTLLIYTAAEVLKRLKELKVWNFLVSLFLVVAGFL
            ATIYLKSHYSFLTDLNILLTLEFLVVYSLLLAVIKKFISVNLFAILISLFILVEMSL
            NASSQMDGIAKEWGFASRSAYSRDILAMESFSTYIGNQFTRTEKLQTQTGNDSMKFNY
            NGISQFSSVRNRSSSSTLDKLKSSGTNLNFYANNSILADSLFGIQYNISDSPIDK
            YGFKDIYQKDNLTLYENQYSLPIAVASQSVYNDVKFNEHTLDNQASFLNQLANVNFDY
            FSPIPYEKTEKIENTNDLISVTSSSNEDAAIQYGIEVPENSQVLSFINLHFSNDKQK
            KVDLIVNGEKKTFTTDNVFSFFNLGYTKEKKTFNIVSFPGNSQVSFESPTFYRLDTK
            TETEAIQKIKEQPVTVSTSKNKVFAYDVQQDTSFFTIPYDKGWSAYQDGKKIEIKQ
            AQTGFMKVDIPKGKGTITLLSFIPNGFITGAICSFTSLLLFGIYNHRRSSKA"
gene        complement(8827. .9141)
            /gene="spr2037"
CDS         complement(8827. .9141)
            /gene="spr2037"
            /codon_start=1
            /transl_table=11
            /label=spr2037
            /product="Conserved hypothetical protein, truncation"
            /protein_id="AAL00839.1"
            /db_xref="GI:15459744"
            /translation="MAGILVTGGVQLIGLTANHEMRIFILILTSLVFMSMVTTLATW
            NSRIGAFFSLILLLQLASSAGTYPLALTNDFFRSINPWLPMSYSVSGLRQTISINKS
            FS"
gene        complement(9169. .9507)
            /gene="spr2038"
CDS         complement(9169. .9507)
            /gene="spr2038"
            /codon_start=1
            /transl_table=11
            /label=spr2038
            /product="Conserved hypothetical protein, truncation"
            /protein_id="AAL00840.1"
            /db_xref="GI:15459745"
            /translation="MADGSRKLAEGGTKLITSGLEDLQTGLASLGQGLGNASDQLKSVS
            TESKNAEILSNPLNLSKTDNDQVPVNGIAIAPYMISVALFFAAISTNMIFAKLPSGRH
            PESRWAWLKS"
gene        complement(9592. .10359)
            /gene="spr2039"
CDS         complement(9592. .10359)
            /gene="spr2039"
            /codon_start=1
            /transl_table=11
            /label=spr2039
            /product="Conserved hypothetical protein, truncation"
            /protein_id="AAL00841.1"
            /db_xref="GI:15459746"
            /translation="MIGLSLIPDLYNIIFLSSMWDPYGQLSDLPVAVVNNDKEASYNG
            NTMAIGKDMVSNLKENKTLDFHFVDEEEGKKGLEDGDYIMVVTLPSDLSEKTTTLSNI
            QSTRAYQSLTSEQQTEISDSVSQNSTDSIQSAQSIVALVQDLQGSLENLQNQSSNLST
```

```
SKSRQATSRKKMLDKIELEEIVPSSRKYPFINFKAEREIGNDLLTVENLTVKIDGETI
LDNISFILRPDKTALIGQNDIQTTALIRAIMGDIDYEGTVKWGVTTSQSYLPKDNSA
DFAGGESILDWLRQFASKEEDNTFLRGFLGRMLFSGDEVNKPVNVLSGGEKVRVMLS
KLMLKSNVLVLDDPTNHLDESISSLNDGLKNFKESIIFASHDHEFIQTLANHIIVL
SKNGVIDRIDETYDEFLENAEVQAKVKELWKD"
```

```
Alignment Scores:
Pred. No.:            11.3
Score:               69.00

Percent Similarity:    50.88%          Conservative:   10
Best Local Similarity: 33.33%          Mismatches:     16
Query Match:           30.40%          Indels:         12
                                       Gaps:            2

US-09-833-017B-2 (1-46) x AE008567 (1-10478)

Qy       2 LysLysThrLeuSerLeuLysAsnPhe-----------------------------LysGlu 13
           ||:|||||      |||                                     |||||
Db    8247 AAAAAGACTTTTACAACTGATAAGTCTTCTCCTTCTTTAATCTAGGATATACTAAAGAG 8306

Qy      14 IleLysThrAspGluLeuGluIleIleGlyGlySerGlySerLeuSer----------- 30
           ||||||           ::|:|||||||:          ::::||
Db    8307 AAAAAAACTTTCAATATCAATGTTAGTTTCCTGGAAATTCACAAGTATCATTTGAATCT 8366

Qy      31 ---ThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 46
              ::||||:|||||                :::::||::|||:||:: |
Db    8367 CCTACCTTCATCGTTTAGATACCAAAACTTTCACCGAGGCAATTCAAAA 8417
```

```
RESULT 11
SPNEU1901
LOCUS       SPNEU1901           232807 bp    DNA        linear   HTG 12-JUL-2001
DEFINITION  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
            in ordered pieces.
ACCESSION   AL449923
VERSION     AL449923.1  GI:11545148
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
ORGANISM    Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 232807)
AUTHORS     Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
            Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
            and Garcia-Bustos,J.F.
TITLE       Annotated draft genomic sequence from a Streptococcus pneumoniae
            type 19F clinical isolate
JOURNAL     Microb. Drug Resist. 7 (2), 99-125 (2001)
MEDLINE     21335329
PUBMED      11442348
REFERENCE   2  (bases 1 to 232807)
AUTHORS     Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
            Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
            Garcia-Bustos,J.F.
TITLE       Direct Submission
JOURNAL     Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
            Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT     * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..232807
                     /organism="Streptococcus pneumoniae"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /clone="G54"
BASE COUNT    70984 a  43538 c  47937 g  70253 t   95 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:            461
Score:               69.00
Percent Similarity:    50.88%          Length:       232807
Best Local Similarity: 33.33%          Matches:          19
Query Match:           30.40%          Conservative:     10
                                       Mismatches:       16
                                       Indels:           12
                                       Gaps:              2

US-09-833-017B-2 (1-46) x SPNEU1901 (1-232807)

Qy       2 LysLysThrLeuSerLeuLysAsnPhe-----------------------------LysGlu 13
```

```
Db  48135  |||||||||   :::   |||   |||        |||||||
           AAAAGACTTTACACAACTGATAAATGTCTTCCTCCTTCTTTAATCTAGGATATACTACTAAAGAG 48194
Qy      14 IleLysThrAspGluLeuGlyLeuIleIleGlyGlySerLeuSer-------- 30
           |||||          |:||  :::      |||:|||            |:|||
Db  48195  AAAAAAACTTTCAATATCAATGTTAGTTTCCCTGGAAATTCACAAGTATCATTTGAATCT 48254
Qy      31 --ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 46
           ||||||:::||||||          :::|::|||:||||:::|||:::|||||
Db  48255  CCTACCCTTCCTATCGTTTAGATACCAAAACTTTCACCGAGGCAATTCAAAAA 48305


RESULT 12
AC107022/c
LOCUS       AC107022        116557 bp    DNA    linear    PRI 27-APR-2002
DEFINITION  Homo sapiens 3 BAC RP11-417H23 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC107022
VERSION     AC107022.4  GI:20334515
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116557)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,R., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payon,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 116557)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (14-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 116557)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 116557)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (27-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 27, 2002 this sequence version replaced gi:18449824.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at, the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
              Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1..116557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-417H23"
     misc_feature    1..1995
                     /note="overlaps bases 190327..192322 of clone AC074043"
                     /function="clone overlap"
     repeat_region   complement(1..77)
                     /rpt_family="THE1C"
     repeat_region   170..561
                     /rpt_family="LTR40b"
     repeat_region   957..1068
                     /rpt_family="MER81"
     repeat_region   complement(1069..1363)
                     /rpt_family="LTR40a"
     STS             1518..1651
                     /standard_name="75005"
     repeat_region   1760..1787
```

```
Query Match:              29.52%     Indels:        0
DB:                       9          Gaps:          0

US-09-833-017B-2 (1-46) x AC107022 (1-116557)

Qy       2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle 21
           |||||| :::|| :::|||  ||| ::::|||   :::|||||||||
Db   61354 AAAAAACACATACACACTCTCAAAGAGTTTGAGAAGGAAACAGACGAATGAGTGGTG 61295

Qy      22 IleIleGlyGlySerLeuSerThrPhePheArgLeuPhe 36
           |||  ::::|||  |::|||:::|||| ||||||
Db   61294 ACTGGAGGGAAAAATGGGGTCATGTCATCATTTTTGTTTGTTT 61250

RESULT 13
ACCESSION   AC009651/c
LOCUS       AC009651             149335 bp   DNA   linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-417H23 map 3, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC009651
VERSION     AC009651.4  GI:9966269
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149335)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-417H23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149335)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,D.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 149335)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,M., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Sep 4, 2000 this sequence version replaced gi:7622330.
```

```
repeat_region    /rpt_family="(A)n"
                 complement(2167..2680)
repeat_region    /rpt_family="MLT1E2"
                 4583..4710
repeat_region    /rpt_family="MIR"
                 4793..4942
repeat_region    /rpt_family="MLT1J2"
                 4968..5152
repeat_region    /rpt_family="MLT1J1"
                 6561..6592
repeat_region    /rpt_family="(TTG)n"
                 6593..6926
repeat_region    /rpt_family="MLT1J2"
                 7417..7731
repeat_region    /rpt_family="LTR16B"
                 7812..7846
repeat_region    /rpt_family="AT_rich"
                 7901..7927
repeat_region    /rpt_family="(TTTG)n"
                 complement(7928..8219)
repeat_region    /rpt_family="AluJo"
                 complement(8424..8753)
repeat_region    /rpt_family="MLT1E"
                 complement(8880..8944)
repeat_region    /rpt_family="MLT1F"
                 8946..8971
repeat_region    /rpt_family="AT_rich"
                 9089..9123
repeat_region    /rpt_family="AT_rich"
                 9221..9411
STS              /standard_name="8620"
                 9317..9374
repeat_region    /rpt_family="(TAGA)n"
                 10733..10771
repeat_region    /rpt_family="(CA)n"
                 complement(10774..10890)
repeat_region    /rpt_family="MLT1J2"
                 complement(10921..11029)
repeat_region    /rpt_family="MLT1J2"
                 complement(11082..11394)
repeat_region    /rpt_family="AluJb"
                 11440..11526
repeat_region    /rpt_family="Charlie4"
                 11717..12040
repeat_region    /rpt_family="LTR16C"
                 13305..13337
repeat_region    /rpt_family="(CAAAA)n"
                 complement(13617..13692)
repeat_region    /rpt_family="MLT1G"
                 complement(13797..14007)
repeat_region    /rpt_family="MLT1G1"
                 14008..14313
repeat_region    /rpt_family="AluJb"
                 complement(14314..14352)
repeat_region    /rpt_family="MLT1G1"
                 14527..15541
repeat_region    /rpt_family="L1M4"
                 15624..16142
repeat_region    /rpt_family="L1M4"
                 16152..16450
repeat_region    /rpt_family="AluSx"
                 16479..18328
repeat_region    /rpt_family="L1ME1"
                 18354..18656
repeat_region    /rpt_family="AluJb"
                 18720..19647
repeat_region    /rpt_family="L1ME1"
```

```
Alignment Scores:
Pred. No.:                        403        Length:              116557
Score:                         67.00        Matches:                 15
Percent Similarity:           62.86%        Conservative:             7
Best Local Similarity:        42.86%        Mismatches:              13
```

. . Wed Nov 13 10:42:10 2002

```
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
            Project Information
    Center project name: L2180
    Center clone name: 417_H_23
            Summary Statistics
    Sequencing vector: M13; M77815; 98% of reads
    Sequencing vector: Plasmid; n/a; %-0.f%% of reads
    2.466699555994080Chemistry: Dye-primer-amersham; 4% of reads
    Chemistry: Dye-terminator Big Dye; 96% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 143822 bases at least Q40
    Consensus quality: 146616 bases at least Q30
    Consensus quality: 147852 bases at least Q20
    Insert size: 182000; agarose-fp
    Insert size: 148935; sum-of-contigs
    Quality covera.
*   NOTE: This is a 'working draft' sequence. It currently
*   consists of 5 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*       1    13133: contig of 13133 bp in length
*   13134 13233: gap of    100 bp
*   13234 18557: contig of 5324 bp in length
*   18558 18657: gap of    100 bp
*   18658 35242: contig of 16585 bp in length
*   35243 35342: gap of    100 bp
*   35343 77529: contig of 42187 bp in length
*   77530 77629: gap of    100 bp
*   77630 149335: contig of 71706 bp in length.
FEATURES             Location/Qualifiers
    source          1..149335
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="3"
                    /map="3"
                    /clone="RP11-417H23"
                    /clone_lib="RPCI-11 Human Male BAC"
    misc_feature    1..13133
                    /note="assembly_fragment"
                    /clone_end:left"
                    vector_side:left"
    misc_feature    13234..18557
                    /note="assembly_fragment"
    misc_feature    18658..35242
                    /note="assembly_fragment"
    misc_feature    35343..77529
                    /note="assembly_fragment"
    misc_feature    77630..149335
                    /note="assembly_fragment"
                    clone_end:T7
                    vector_side:right"
BASE COUNT   46161 a  28722 c  28526 g  45525 t     401 others
ORIGIN

Alignment Scores:
Pred. NO.:                  542           Length:             149335
Score:                      67.00         Matches:            15
Percent Similarity:         62.86%        Conservative:       7
Best Local Similarity:      42.86%        Mismatches:         13
Query Match:                29.52%        Indels:             0
DB:                         2             Gaps:               0
```

```
US-09-833-017B-2 (1-46) x AC009651 (1-149335)

Qy      2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrGluLeuGluIle 21
          |||||     |:::   |||      ||:::||| :::||||||  ::||||||||   :::
Db  35536 AAAAAACACATACAACTCTCAAAGAGTTTGAAGAAGGAAACAGACGAATGAGTGGTG 35477

Qy     22 IleIleGlyGlySerGlyLeuSerThrPhePheArgLeuPhe 36
          |||      |:::   ||||||      ||:::||    |:||||
Db  35476 ACTCGAGGAAAATGGGGTCATGTCATCATTTTTGTTTGTTT 35432

RESULT 14
AC067831/c
LOCUS       AC067831              174028 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-710K1 map 3, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC067831
VERSION     AC067831.2  GI:8099890
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-710K1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda.F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,Al., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar.T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,CL., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy.P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 174028)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda.F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel.Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 28, 2000 this sequence version replaced gi:7651882.
          All repeats were identified using RepeatMasker.
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ------ Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          -------- Project Information
          Center project name: L9432
          Center clone name: 710_K_1
          -------- Summary Statistics
          Sequencing vector: M13; M77815; 100% of reads
          Chemistry: Dye-terminator Big Dye; version 0.960731
          Assembly program: Phrap; version 0.960731
          Consensus quality: 163664 bases at least Q40
          Consensus quality: 169434 bases at least Q30
          Consensus quality: 171495 bases at least Q20
          Insert size: 176000; agarose-fp
          Insert size: 172528; sum-of-contigs
          Quality coverage: 4.4 in Q20 bases; agarose-fp
          Quality coverage: 4.5 in Q20 bases; sum-of-contigs
          --------
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 16 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        * -------
        *      1    1381: contig of 1381 bp in length
        * 1382   1481: gap of   100 bp in length
        * 1482   2059: contig of  578 bp in length
        * 2060   2159: gap of   100 bp
        * 2160   3484: contig of 1325 bp in length
        * 3485   3584: gap of   100 bp
        * 3585   5811: contig of 2227 bp in length
        * 5812   5911: gap of   100 bp
        * 5912   9908: contig of 3997 bp in length
        * 9909  10008: gap of   100 bp
        * 10009  17481: contig of 7473 bp in length
        * 17482  17581: gap of   100 bp
        * 17582  26280: contig of 8699 bp in length
        * 26281  26380: gap of   100 bp
        * 26381  34007: contig of 7627 bp in length
        * 34008  34107: gap of   100 bp
        * 34108  46064: contig of 11957 bp in length
        * 46065  46164: gap of   100 bp
        * 46165  57569: contig of 11405 bp in length
        * 57570  57669: gap of   100 bp
        * 57670  71968: contig of 14299 bp in length
        * 71969  72068: gap of   100 bp
        * 72069  85308: contig of 13240 bp in length
        * 85309  85408: gap of   100 bp
        * 85409 104044: contig of 18636 bp in length
        * 104045 104144: gap of   100 bp
        * 104145 128575: contig of 24431 bp in length
        * 128576 128675: gap of   100 bp
        * 128676 148603: contig of 19928 bp in length
        * 148604 148703: gap of   100 bp
        * 148704 174028: contig of 25325 bp in length.

FEATURES     Location/Qualifiers
     source    1..174028
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="3"
               /map="3"
               /clone="RP11-710K1"
               /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1381
                     /note="assembly_fragment"
     misc_feature    1482..2059
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    2160..3484
                     /note="assembly_fragment"
     misc_feature    3585..5811
                     /note="assembly_fragment"
     misc_feature    5912..9908
                     /note="assembly_fragment"
     misc_feature    10009..17481
                     /note="assembly_fragment"
     misc_feature    17582..26280
                     /note="assembly_fragment"
     misc_feature    26381..34007
                     /note="assembly_fragment"
     misc_feature    34108..46064
                     /note="assembly_fragment"
     misc_feature    46165..57569
                     /note="assembly_fragment"
     misc_feature    57670..71968
                     /note="assembly_fragment"
     misc_feature    72069..85308
                     /note="assembly_fragment"
     misc_feature    85409..104044
                     /note="assembly_fragment"
     misc_feature    104145..128575
                     /note="assembly_fragment"
     misc_feature    128676..148603
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    148704..174028
                     /note="assembly_fragment"
BASE COUNT    53885 a   34369 c   32461 g   51811 t    1502 others
ORIGIN

Alignment Scores:
Pred. No.:                651
Score:                    67.00
Percent Similarity:       62.86%          Length:           174028
Best Local Similarity:    42.86%          Matches:              15
Query Match:              29.52%          Conservative:          7
Db:                       2               Mismatches:           13
                                          Indels:                0
                                          Gaps:                  0

US-09-833-017B-2 (1-46) x AC067831 (1-174028)

Qy    2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle 21
          |||||  :::  ||   :::::::::||  :::||||||||||
Db 8172 AAAAACACCATACAACTCTCAAAGAGTTTGAAGAAGGGAAACACGACGAATGAGTGGTG 8113

Qy   22 IleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPhe 36
          |||  :::||| ::::|| :::||||| ||||||
Db 8112 ACTGGAGGGAAAATGGGGTCATGTCATCATTTTTGTTTTGTTT 8068

RESULT 15
AC063944/C
LOCUS       AC063944     180389 bp    DNA      linear    PRI 28-MAR-2002
DEFINITION  Homo sapiens 3 BAC RP11-446H18 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC063944
VERSION     AC063944.25   GI:19774261
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180389)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
```

Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,R., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,R.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,R., Luna,R., Martindale,A.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,J., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
Gibbs,R.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 180389)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (22-APR-2000) Human Genome Sequencing Center, Department
    of Molecular and Human Genetics, Baylor College of Medicine, One
    Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 180389)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
    of Molecular and Human Genetics, Baylor College of Medicine, One
    Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 180389)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (28-MAR-2002) Human Genome Sequencing Center, Department
    of Molecular and Human Genetics, Baylor College of Medicine, One
    Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Mar 28, 2002 this sequence version replaced gi:19718609.
    INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
    gc-help@bcm.tmc.edu

    CLONE LENGTH: This sequence does not necessarily represent the
    entire insert of this clone. Overlapping regions of clones are only
    sequenced and submitted once, so the sequence for the remainder of
    the insert may be found in the record for the adjacent clones.
    Overlapping clones are noted at the beginning and end of the
    Features listing.

    ANNOTATION OF FEATURES:
    STSs are identified using ePCR (Genome Res. 7:541-550) searches
    of a local database that includes entries from dbSTS, GDB, and

    local mapping efforts.
    Repeats are identified using RepeatMasker (A. Smit and P. Green,
    unpublished.) for Human and Mouse sequences.
    Genes and Region of sequence similarity are identified by BLAST
    (Nuc. Acids Res 25.3389-3402) similarity (expect < 1e-34) to the
    EST and cDNA sequences. Genes demonstrate at least two exons
    flanked by consensus splice sites that maintained sequence
    continuity across the splice junctions. Sequences that are not
    identical matches are annotated as similar.

    SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
    standard of double strand coverage with a minimum of 2 clones and 2
    reads with no ambiguities or 2 chemistries with a minimum of 2
    clones and 3 reads with no ambiguities. If the sequence quality for
    a region does not meet this standard, it will be indicated in the
    annotation as Low Coverage.

    QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
    standards - estimated error rate less than 1 per 10,000 bases.
    Reports of lowest quality individual bases and measures of base
    quality are listed below. Description of the metrics can be found
    at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
    html.

FEATURES   Location/Qualifiers
    source   1..180389
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="3"
      /clone="RP11-446H18"
    misc_feature   1..2015
      /note="overlaps bases 190308. .192322 of clone AC074043"
      /function="clone overlap"
    repeat_region   complement(1..96)
      /rpt_family="L1M4"
    repeat_region   189..580
      /rpt_family="THE1C"
    repeat_region   complement(752..975)
      /rpt_family="LTR40b"
    repeat_region   976..1087
      /rpt_family="MER81"
    repeat_region   complement(1088..1382)
      /rpt_family="LTR40a"
    STS   1537..1670
      /standard_name="75005"
    repeat_region   1779..1807
      /rpt_family="(A)n"
    repeat_region   complement(2187..2700)
      /rpt_family="MLT1E2"
    repeat_region   4696..4823
      /rpt_family="MIR"
    repeat_region   4906..5055
      /rpt_family="MLT1J1"
    repeat_region   5081..5265
      /rpt_family="MLT1J1"
    repeat_region   6184..6204
      /rpt_family="AT_rich"
    repeat_region   6675..6706
      /rpt_family="AT_rich"
    repeat_region   6707..7040
      /rpt_family="(TTG)n"
    repeat_region   7531..7845
      /rpt_family="MLT1J2"
    repeat_region   7926..7960
      /rpt_family="LTR16B"
    repeat_region   8015..8041
      /rpt_family="AT_rich"
    repeat_region   complement(8042..8333)
      /rpt_family="(TTTG)n"
    repeat_region   complement(8538..8867)
      /rpt_family="AluJo"
    repeat_region   complement(8994..9058)
      /rpt_family="MLT1E"

QUALSTAT-REPORT.

```
repeat_region    /rpt_family="MLT1F"
                 9060..9085
repeat_region    /rpt_family="AT_rich"
                 9203..9237
STS              /rpt_family="AT_rich"
                 9335..9525
                 /standard_name="8620"
repeat_region    9431..9488
                 /rpt_family="(TAGA)n"
repeat_region    10847..10885
                 /rpt_family="(CA)n"
repeat_region    complement(10888..11004)
                 /rpt_family="MLT1J2"
repeat_region    complement(11035..11143)
                 /rpt_family="MLT1J2"
repeat_region    complement(11196..11508)
                 /rpt_family="AluJb"
repeat_region    11554..11640
                 /rpt_family="Charlie4"
repeat_region    11831..12154
                 /rpt_family="LTR16C"
repeat_region    13419..13451
                 /rpt_family="(CAAAA)n"
repeat_region    complement(13731..13806)
                 /rpt_family="MLT1G"
repeat_region    complement(13911..14121)
                 /rpt_family="MLT1G1"
repeat_region    14122..14427
                 /rpt_family="AluJb"
repeat_region    complement(14428..14466)
                 /rpt_family="MLT1G1"
repeat_region    14641..15655
                 /rpt_family="L1M4"
repeat_region    15738..16256
                 /rpt_family="L1M4"
repeat_region    16266..16564
                 /rpt_family="AluSx"
repeat_region    16593..18442
                 /rpt_family="L1ME1"
repeat_region    18468..18770
                 /rpt_family="AluJb"


Alignment Scores:
Pred. No.:                    680.00              Length:          180389
Score:                        67.00              Matches:         15
Percent Similarity:           62.86%             Conservative:    7
Best Local Similarity:        42.86%             Mismatches:      13
Query Match:                  29.52%             Indels:          0
DB:                           9                  Gaps:            0

US-09-833-017B-2 (1-46) x AC063944 (1-180389)

Qy     2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle 21
         ||||||   :::  |||    ||::::||||     ::::||||||||||        :::
Db 61451 AAAAAACACATACAAACTCTCAAAGAGTTTGAAGAAAGGGAAACAAGACGAATGAGTGGTG 61392

Qy    22 IleIleGlyGlySerGlySerLeuSerThrPhePhePheArgLeuPhe 36
         |||  ::::||||  ::::||||:::::||:||||    ||||||
Db 61391 ACTGGAGGGAAAAATGGGGTCATGTCATCATTTTTGTTTTGTTT 61347


Search completed: November  8, 2002, 22:56:29
Job time : 1885.72 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2002, 20:29:27 ; Search time 229.313 Seconds
                                           (without alignments)
                                           451.748 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 227
Sequence: 1 MKKTLSLKNDFKEIKTDELE........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext   0.5
               Ygapop 10.0 , Ygapext   0.5
               Fgapop   6.0 , Fgapext   7.0
               Delop   6.0 , Delext   7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn_1/USPTO_spool/US09833017/runat_05112002_105348_4824/app_query.fasta_1.398
-DB=N_Geneseq_101002 -QFMT=fastap -SUFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017_@CGN_1_1_125_@runat_05112002_105348_4824 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=4 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_101002:*
         1:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
         2:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
         3:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
         4:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
         5:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
         6:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
         7:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
         8:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
         9:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
        10:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
        11:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
        12:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
        13:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
        14:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
        15:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
        16:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
        17:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
        18:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
        19:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
        20:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
        21:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
        22:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
        23:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
        24:    /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB		ID	Description
1	227	100.0	141	24		AAD32791	Streptococcus muta
2	227	100.0	141	24		AAD32898	Streptococcus muta
3	227	100.0	2557	24		AAD32800	Streptococcus muta
4	227	100.0	2557	24		AAD32893	Streptococcus muta
5	104	45.8	63	24		AAD32790	Streptococcus muta
6	104	45.8	63	24		AAD32884	Streptococcus muta
7	69	30.4	8411	19		AAV52149	Streptococcus pneu
8	59	26.0	681	24	c	ABQ15020	Oligonucleotide fo
9	59	26.0	681	24	c	ABQ15021	Oligonucleotide fo
10	59	26.0	1127	22	c	AAH87647	Peppermint plant o
11	59	26.0	1481	22	c	AAF85076	Nucleotide sequenc
12	58.5	25.8	416	24		ABK77550	Bacillus clausii g
13	58.5	25.8	885	24		ABQ46476	Oligonucleotide fo
14	58.5	25.8	885	24		ABQ46477	Oligonucleotide fo
15	58.5	25.8	1530	24	c	ABN70018	Streptococcus poly
16	58.5	25.8	1533	22	c	AAS06359	Streptococcus pyog
17	58.5	25.8	37661	22	c	AAF90334	Human neurokinin B
18	58	25.6	691	18	c	AAT84204	DNA encoding a Sta
19	58	25.6	4702	18	c	AAV74579	Staphylococcus aur
20	58	25.6	6389	24	c	ABL33960	Human immune syste
21	57	25.1	942	24	c	ABK74804	Bacillus lichenifo
22	57	25.1	48667	24	c	ABK51628	Human transporter
23	56.5	24.9	843	24		ABK75880	Bacillus lichenifo
24	56.5	24.9	1061	11	c	AAQ03182	Glyceraldehyde-pho
25	56.5	24.9	13884	20		AAX13278	Enterococcus faeca
26	56	24.7	2820	23		AAS53063	Enterococcus faeca
27	56	24.7	5480	20		AAX13069	Enterococcus faeca
28	56	24.7	6467	24		ABN80169	Human chemically m
29	56	24.7	465237	24	c	ABQ87681	Human oestrogen re
30	56	24.7	465237	24	c	ABA90193	Human oestrogen re
31	55.5	24.4	1251	24		ABQ67878	Listeria monocytog
32	55.5	24.4	1662	17		AAT18203	Newcastle disease
33	55.5	24.4	3825	8		AAN70261	Sequence of Newcas
34	55.5	24.4	4177	15		AAQ68943	SfiI fragment cont
35	55.5	24.4	4177	15		AAQ70570	SfiI fragment cont
36	55.5	24.4	4177	18		AAT48510	SfiI fragment enco
37	55.5	24.4	4177	20		AAX81147	Seq ID No: 12 of U
38	55.5	24.4	4177	21		AAC67862	SfiI fragment comp
39	55.5	24.4	4177	21		AAZ49295	cDNA encoding NDV
40	55.5	24.4	15186	21		AAZ44617	Newcastle disease
41	55	24.2	367	22		AAI91483	Human polynucleoti
42	55	24.2	1752	22		AAH52982	S. epidermidis ope
43	55	24.2	1824	24		ABN79814	Fungal zBC gene se
44	55	24.2	3033	24		ABN92878	staphylococcus epi
45	55	24.2	3055	22		AAH54711	S. epidermidis gen

ALIGNMENTS

RESULT 1
AAD32791
ID   AAD32791 standard; DNA; 141 BP.
XX
AC   AAD32791;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans comC gene.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; gene; comC gene; ds.
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..141

```
FT
FT                   /*tag=  a
FT                   /product= "S. mutans ComC protein"
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCY, Cvitkovitch DG, Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB; AAE20494.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis -
XX
PS   Disclosure; Fig 2; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. Vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans ComC CSP precursor gene.
XX
SQ   Sequence 141 BP; 58 A; 19 C; 25 G; 39 T; 0 other;

Alignment Scores:
Pred. No.:              2.59e-28       Length:        141
Score:                  227.00         Matches:       46
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     24             Gaps:          0

US-09-833-017B-2 (1-46) x AAD32791 (1-141)

Qy    1 MetLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAAAAAAACACTTCATTAAAAAATGACTTTAAAGAAATTAAGACTGAATTAGAG 60

Qy   21 IleIleIleGlyGlySerGlySerThrPhePheArgLeuPheAsnArgSerPhe 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTATCATTGGCGGAGGCGGAAGCGGAAGCCTATCAACATTTTTCGCGGCTGTTTT 120

Qy   41 ThrGlnAlaLeuGlyLys 46
          ||||||||||||||||||
Db  121 ACACAAGCTTTGGGAAAA 138

RESULT 2
AAD32898
ID   AAD32898 standard; DNA; 141 BP.
XX
AC   AAD32898;
```

```
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans ComC gene.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
KW   therapy; caries; endocarditis; microbial biofilm; infection; gene;
KW   vaccine; antibacterial; ComC gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..141
FT                   /*tag=  a
FT                   /product= "S. mutans ComC protein"
XX
PN   CA2332733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/) LI Y H.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LAUP/) LAU P C.
XX
PI   Li YH, Cvitkovitch DG, Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB; AAE20594.
XX
PT   Novel compound that competitively inhibits binding of competence signal
PT   peptide to Streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis -
XX
PS   Disclosure; Fig 2; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans ComC gene encoding a
CC   CSP precursor protein.
XX
SQ   Sequence 141 BP; 58 A; 19 C; 25 G; 39 T; 0 other;

Alignment Scores:
Pred. No.:              2.59e-28       Length:        141
Score:                  227.00         Matches:       46
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     24             Gaps:          0

US-09-833-017B-2 (1-46) x AAD32898 (1-141)
```

```
Qy   1 MetLysThrLeuSerLeuLysAsnAspPheLysIleLysThrAspGluLeuGlu 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60
Qy  21 IleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPhe 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTT 120
Qy  41 ThrGlnAlaLeuGlyLys 46
       ||||||||||||
Db 121 ACACAAGCTTTGGGAAAA 138

RESULT 3
ID   AAD32800 standard; DNA; 2557 BP.
XX
AC   AAD32800;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans comCDE gene local region.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; comCDE gene; gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (25..168)
FT                   /tag= a
FT                   /product= "Protein #1 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             complement (452..499)
FT                   /tag= b
FT                   /product= "Peptide #1 encoded by S. mutans comCDE gene
FT                   local region"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
FT   CDS             648..758
FT                   /tag= c
FT                   /product= "Peptide #2 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             953..1081
FT                   /tag= d
FT                   /product= "Protein #2 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             complement (1366..1449)
FT                   /tag= e
FT                   /product= "Peptide #3 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             1855..1959
FT                   /tag= f
FT                   /product= "Peptide #4 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             complement (1896..2072)
FT                   /tag= g
FT                   /product= "Protein #3 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             2182..2424
FT                   /tag= h
FT                   /product= "Protein #4 encoded by S. mutans comCDE gene
FT                   local region"
FT   CDS             2384..2488
FT                   /tag= i
FT                   /product= "Peptide #5 encoded by S. mutans comCDE gene
FT                   local region"
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
```

```
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCY, Cvitkovitch DG, Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB; AAE20607, AAE20608, AAE20609, AAE20610, AAE20611, AAE20612,
DR   AAE20613, AAE20614, AAE20615.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis
XX
PS   Disclosure; Fig 9; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans comCDE gene local region.
XX
SQ   Sequence 2557 BP; 891 A; 415 C; 426 G; 825 T; 0 other;

Alignment Scores:
Pred. No.:                        1.27e-26        Length:          2557
Score:                            227.00          Matches:         46
Percent Similarity:               100.00%         Conservative:    0
Best Local Similarity:            100.00%         Mismatches:      0
Query Match:                      100.00%         Indels:          0
DB:                               24              Gaps:            0

US-09-833-017B-2 (1-46) x AAD32800 (1-2557)

Qy   1 MetLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 20
Qy  21 IleIleIleGlyGlySerGlySerLeuSerGlyLeuPheAsnArgSerPhe 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 160
Qy  21 IleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPhe 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTT 220
Qy  41 ThrGlnAlaLeuGlyLys 46
       ||||||||||||
Db 221 ACACAAGCTTTGGGAAAA 238

RESULT 4
ID   AAD32893 standard; DNA; 2557 BP.
XX
AC   AAD32893;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans comCDE gene local region.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
```

KW   therapy; caries; endocarditis; microbial biofilm; infection; gene;
KW   vaccine; antibacterial; comCDE gene local region; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   complement (25..168)
FT   /*tag= a
FT   /product= "Protein #1 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   101..241
FT   /*tag= b
FT   /product= "S. mutans ComC protein"
FT   CDS   complement (383..1708)
FT   /*tag= c
FT   /product= "S. mutans ComD protein"
FT   CDS   complement (452..499)
FT   /*tag= d
FT   /product= "Peptide #1 encoded by S. mutans ComCDE
FT   gene local region"
FT   /note= "CDS does not include start and stop codon"
FT   /partial
FT   CDS   648..758
FT   /*tag= e
FT   /product= "Peptide #2 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   953..1081
FT   /*tag= f
FT   /product= "Protein #2 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   complement (1366..1449)
FT   /*tag= g
FT   /product= "Peptide #3 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   (1705..2457)
FT   /*tag= h
FT   /product= "S. mutans ComE protein"
FT   CDS   1855..1959
FT   /*tag= i
FT   /product= "Peptide #4 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   complement (1896..2072)
FT   /*tag= j
FT   /product= "Protein #3 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   complement (2182..2424)
FT   /*tag= k
FT   /product= "Protein #4 encoded by S. mutans ComCDE
FT   gene local region"
FT   CDS   2384..2488
FT   /*tag= l
FT   /product= "Peptide #5 encoded by S. mutans ComCDE
FT   gene local region"
XX
PN   CA2332733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/) LI Y H.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LAUP/) LAU P C.
XX
PI   Li YH, Cvitkovitch DG, Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB; AAE20617, AAE20618, AAE20619, AAE20620, AAE20621, AAE20622,
DR   AAE20623, AAE20624, AAE20625, AAE20594, AAE20585, AAE20586.
XX
PT   Novel compound that competitively inhibits binding of competence signal

PT   peptide to Streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis -
XX
PS   Disclosure; Fig 9; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans ComCDE gene local
CC   region.
XX
SQ   Sequence 2557 BP; 891 A; 415 C; 426 G; 825 T; 0 other;

Alignment Scores:
Pred. No.:   2557
Score: 1.27e-26   Matches: 46
Percent Similarity: 227.00   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 24   Gaps: 0

US-09-833-017B-2 (1-46) x AAD32893 (1-2557)

Qy   1   MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101   ATGAAAAAACACTATCATTAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 160

Qy   21   IleIleIleGlyGlySerGlySerLeuThrPhePheArgLeuPheAsnArgSerPhe 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161   ATTATCATTGGCGGAAGCGGAAGCCCTATCAACACATTTTCCGGCTGTTTAACAGAAGTTTT 220

Qy   41   ThrGlnAlaLeuGlyLys 46
          ||||||||||||||||||
Db 221   ACACAAGCTTTGGGAAAA 238

RESULT 5
AAD32790
ID   AAD32790 standard; DNA; 63 BP.
XX
AC   AAD32790;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans competence signal peptide (CSP) DNA.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..63
FT   /*tag= a
FT   /product= "S. mutans competence signal peptide (CSP)"
FT   /note= "CDS does not include start and stop codon"

```
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
KW   therapy; caries; endocarditis; microbial biofilm; infection; gene;
KW   vaccine; antibacterial; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..63
FT                   /*tag= a
FT                   /product= "S. mutans competence signal peptide"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   CA2332733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/) LI Y H.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LAUP/) LAU P C.
XX
PI   Li YH, Cvitkovitch DG, Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB; AAE20584.
XX
PT   Novel compound that competitively inhibits binding of competence signal
PT   peptide to Streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis -
XX
PS   Claim 8; Fig 2; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans CSP DNA.
XX
SQ   Sequence 63 BP; 20 A; 12 C; 13 G; 18 T; 0 other;

Alignment Scores:
Pred. No.:              1.01e-08         Length:        63
Score:                  104.00           Matches:       21
Percent Similarity:     100.00%          Conservative:  0
Best Local Similarity:  100.00%          Mismatches:    0
Query Match:            45.81%           Indels:        0
DB:                     24               Gaps:          0

US-09-833-017B-2 (1-46) x AAD32884 (1-63)

Qy   26 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 45
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   26 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 45
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAGAGTTTTACACAAGCTTTGGGA 60
```

```
FT                   /partial
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCy, Cvitkovitch DG, Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB; AAE20493.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis -
XX
PS   Claim 8; Fig 2; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. Vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans CSP DNA.
XX
SQ   Sequence 63 BP; 20 A; 12 C; 13 G; 18 T; 0 other;

Alignment Scores:
Pred. No.:              1.01e-08         Length:        63
Score:                  104.00           Matches:       21
Percent Similarity:     100.00%          Conservative:  0
Best Local Similarity:  100.00%          Mismatches:    0
Query Match:            45.81%           Indels:        0
DB:                     24               Gaps:          0

US-09-833-017B-2 (1-46) x AAD32790 (1-63)

Qy   26 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 45
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAGAGTTTTACACAAGCTTTGGGA 60

Qy   46 Lys 46
         |||
Db   61 AAA 63

RESULT 6
AAD32884
ID   AAD32884 standard; DNA; 63 BP.
XX
AC   AAD32884;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans competence signal peptide (CSP) DNA.
XX
```

```
Qy      46 Lys 46                                              19          Gaps:       2
           |||
Db      61 AAA 63

RESULT 7
AAV52149/c
ID   AAV52149 standard; DNA; 8411 BP.
XX
AC   AAV52149;
XX
DT   23-OCT-1998 (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:16.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997; 97WO-US19588.
XX
PR   31-OCT-1996; 96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI   Kunsch CA, Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 239-244; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 8411 BP; 2741 A; 1386 C; 1639 G; 2645 T; 0 other;

Alignment Scores:
Pred. No.:              3.67              Length:              8411
Score:                 69.00              Matches:               19
Percent Similarity:    50.88%             Conservative:          10
Best Local Similarity: 33.33%             Mismatches:            16
Query Match:           30.40%             Indels:                12
```

```
DB:                              19          Gaps:       2

US-09-833-017B-2 (1-46) x AAV52149 (1-8411)

Qy     2 LysLysThrLeuLeuSerLeuLysAsnAspPhe------------------LysGlu 13
         ||  |||||   ||||   |||||
Db  3566 AAAAAGACTTTTACAACTGATAATGTCTTCTTCCTTCTTAATCTAGGATATACTAAAGAG 3507

Qy    14 IleLysThrAspGluLeuGluIleIleIleGlyGlySerGlySerLeuSer-------- 30
         ||||||       ::  ::  ::::        ::::        ||
Db  3506 AAAAAAACTTTCAATATCAATGTTAGTTTCCTGGAATTCACAAGTATCATTTGAATCT 3447

Qy    31 --ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 46
           |||||::::|||||      :::::::|||||::::|||::: |||
Db  3446 CCTACCTTCTATCGTTTAGATACCAAAACTTTCACCGAGGCAATTCAAAAA 3396
```

```
RESULT 8
ABQ15020
ID   ABQ15020 standard; DNA; 681 BP.
XX
AC   ABQ15020;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 1611.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
```

CC   the disclosure of the invention.
XX
SQ   Sequence 681 BP; 69 A; 101 C; 283 G; 228 T; 0 other;

Alignment Scores:
Pred. No.:   5.38   Length: 681
Score: 59.00   Matches: 11
Percent Similarity: 60.00%   Conservative: 4
Best Local Similarity: 44.00%   Mismatches: 10
Query Match: 25.99%   Indels: 0
DB: Gaps: 0

US-09-833-017B-2 (1-46) x ABQ15020 (1-681)

Qy   21 IleIleIleGlyGlySerThrPheArgLeuPheAsnArgSerPhe 40
        :::::: |||||| |||||| |||.:: ||||||
Db 255 TTGTTGTTCGGCGGCGGCGGTCGTAGTTCGGGTTTTTTCGGGCGTTTTCGTCGTTC 314

Qy   41 ThrGlnAlaLeuGly 45
        :::
Db 315 GCGCGCGGTTATGGA 329

RESULT 9
ABQ15021/c
ID   ABQ15021 standard; DNA; 681 BP.
XX
AC   ABQ15021;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 1612.
XX
KW   Human; cytosine methylation: 5'-CpG-3'; uracil; cytosine: diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (1) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory

CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 681 BP; 228 A; 283 C; 101 G; 69 T; 0 other;

Alignment Scores:
Pred. No.:   5.38   Length: 681
Score: 59.00   Matches: 11
Percent Similarity: 60.00%   Conservative: 4
Best Local Similarity: 44.00%   Mismatches: 10
Query Match: 25.99%   Indels: 0
DB: Gaps: 0

US-09-833-017B-2 (1-46) x ABQ15021 (1-681)

Qy   21 IleIleIleGlyGlySerLeuSerThrPhePheAsnArgSerPhe 40
        |||||| ||| ||| |||:::: |||||
Db 427 TTGTTGTTCGGCGGCGGCGGGTCGTAGTTCGGGTTTTTTCGGGCGGTTTTCGTCGTCGTTC 368

Qy   41 ThrGlnAlaLeuGly 45
        :::
Db 367 GCGCGCGGTTATGGA 353

RESULT 10
AAH87647/C
ID   AAH87647 standard; cDNA; 1127 BP.
XX
AC   AAH87647;
XX
DT   25-SEP-2001 (first entry)
XX
DE   Peppermint plant oil gland expressed cDNA 3.
XX
KW   Peppermint; plant oil gland cell; terpenoid essential oil; resin;
KW   genetic mapping; antisense suppression; recombinant expression; ss.
XX
OS   Mentha x piperita.
XX
PN   WO20153319-A1.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US02567.
XX
PR   20-JAN-2000; 2000US-0177264.
XX
PA   (CROT/) CROTEAU R B.
PA   (LANG/) LANGE B M.
PA   (WILD/) WILDUNG M R.
XX
PI   Croteau RB, Lange BM, Wildung MR;
XX
DR   WPI; 2001-488706/53.
XX
PT   New nucleic acid molecules corresponding to mRNA molecules expressed in
PT   peppermint oil glands for enhancing expression of plant oil gland cell
PT   proteins -
XX
PS   Claim 1; Page 77; 251pp; English.
XX
CC   The invention relates to nucleic acid molecules (AAH87645-AAH88116) that
CC   correspond to all or part of a mRNA molecule expressed in plant oil
CC   gland cells, especially peppermint and plant oil glands that produce
CC   terpenoid essential oils and resins. The nucleic acids are useful for
CC   genetically mapping a plant genome for genes expressed in plant oil
CC   gland cells and to suppress (for example by antisense suppression) or
CC   enhance their expression (for example by genetically transforming a
CC   plant cell with a replicable expression vector that expresses one or more

CC   proteins naturally expressed in plant oil gland cells). The nucleic acids
CC   are also useful for recombinant expression of plant oil gland proteins
CC   required for terpenoid essential oil and/or resin production in bacterial
CC   and/or yeast cells.
XX
SQ   Sequence 1127 BP; 308 A; 244 C; 247 G; 327 T; 1 other;

Alignment Scores:
Pred. No.:   Length: 1127
Score: 10.6   Matches: 17
Percent Similarity: 59.00   Conservative: 6
Best Local Similarity: 56.10%   Mismatches: 14
Query Match: 41.46%   Indels: 4
DB: 25.99%   Gaps: 1
                     22

US-09-833-017B-2 (1-46) x AAH87647 (1-1127)

Qy   1 MetLysLysThrLeuLysSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
         ::::||||||||||| ::: ::| ||| ||||||
Db   334 CTTAAGAAAACATTAATCTTGCAAGGCGAA----------AAAAGAGTGAGTCTTGAA 287

Qy   21 IleIleIleGlyGlySerGlyLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
         ||| ||| ::: || ||||||||:::||||||:::
Db   286 ATCCCAGCCTCCCTATCCACTTCATCAGCTAATTGTTTAATTTATTCTCAAGATCATAT 227

Qy   41 Thr 41
         |||
Db   226 ACC 224

RESULT 11
AAF85076/c
ID   AAF85076 standard; DNA; 1481 BP.
XX
AC   AAF85076;
XX
DT   09-JUL-2001 (first entry)
XX
DE   Nucleotide sequence of an isopentenyl monophosphate kinase.
XX
KW   Isopentenyl monophosphate kinase; IPK; isoprenoid biosynthesis; pigment;
KW   vitamin; essential oil; pathogen resistance; antibiotic; herbicide;
KW   antimalarial; ss.
XX
OS   Mentha piperita.
XX
FH   Key   Location/Qualifiers
FT   CDS   3..1220
FT   /tag= a
FT   /product= "isopentenyl monophosphatase kinase"
XX
PN   WO200132907-A1.
XX
PD   10-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-US30289.
XX
PR   04-NOV-1999; 99US-0434774.
XX
PA   (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI   Croteau RB, Lange BM;
XX
DR   WPI; 2001-308747/32.
DR   P-PSDB; AAB68285.
XX
PT   New nucleic acid encoding isopentenyl monophosphate kinase, useful e.g.
PT   for making transgenic plants with increased synthesis of isoprenoids,
PT   e.g. essential oils -
XX
PS   Claim 12; Page 42-44; 62pp; English.
XX
CC   The present sequence encodes an isopentenyl monophosphate kinase (IPK)
CC   polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The

---

CC   IPK polynucleotide is useful recombinant production of IPK, as a source
CC   of probes, primers and antisense sequences and for increasing/reducing
CC   expression levels of IPK in cells, particularly of essential oil
CC   plants, so as to increase flow through the isoprenoid biosynthesis
CC   pathway, resulting in increased production of e.g. pigments, vitamins
CC   and essential oils, also to increase resistance to pests and pathogens
CC   and to improve plant strength. Mutant forms of the IPK polynucleotide
CC   can be used to express forms of IPK that are resistant to IPK-targeted
CC   herbicides, and recombinant IPK can be used to screen for antibiotics,
CC   herbicides and antimalarial agents directed against IPK.
XX
SQ   Sequence 1481 BP; 413 A; 320 C; 302 G; 446 T; 0 other;

Alignment Scores:
Pred. No.:   Length: 1481
Score: 15.3   Matches: 17
Percent Similarity: 59.00   Conservative: 6
Best Local Similarity: 56.10%   Mismatches: 14
Query Match: 41.46%   Indels: 4
DB: 25.99%   Gaps: 1
                     22

US-09-833-017B-2 (1-46) x AAF85076 (1-1481)

Qy   1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
         :::||||||||||| |||::   :::
Db   334 CTTAAGAAAACATTAATCTTGCAAGGCGAA----------AAAAGAGTGAGTCTTGAA 287

Qy   21 IleIleIleGlyGlySerGlyLeuSerThrPheArgLeuPheAsnArgSerPhe 40
         ||| ||| ::: ||| |||::||||||:::
Db   286 ATCCCAGCCTCCCTATCCACTTCATCAGCTAATTGTTTAATTTATTCTCAAGATCATAT 227

Qy   41 Thr 41
         |||
Db   226 ACC 224

RESULT 12
ABK77550
ID   ABK77550 standard; DNA; 416 BP.
XX
AC   ABK77550;
XX
DT   13-AUG-2002 (first entry)
XX
DE   Bacillus clausii genomic sequence tag (GST) #393.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus clausii.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO) NOVOZYMES BIOTECH INC.
PA   (NOVO) NOVOZYMES AS.
XX
PI   Berka R, Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array -
XX
PS   Claim 11; SEQ ID NO 4841; 200pp; English.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 885 BP; 127 A; 99 C; 327 G; 332 T; 0 other;

Alignment Scores:
Pred. NO.:   9.23   Length: 885
Score: 58.50   Matches: 15
Percent Similarity: 41.67%   Conservative: 10
Best Local Similarity: 25.00%   Mismatches: 16
Query Match: 25.77%   Indels: 19
                          24   Gaps: 1

US-09-833-017B-2 (1-46) x ABQ46476 (1-885)

Qy   5 LeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle--------- 21
           ||||| ::: |||||:|:::::
Db   89 TTGGCGTTTCGTTTCGTTTTCGCGAGGTTCGCGGCGGTTCGGCGGAGGTGGGAGGGTTGG 148

Qy   22 ------------------------------------IleIleGlyGly 25
                                                  ::: |||
Db   149 AGGCGAGTTTCGACGCGAGAGATTGGCCGGATTGGTTGATTGGTAGACGTTGGCGGGA 208

Qy   26 SerGlySerLeuSerThrPhePheAsnArgSerPheThrGlnAlaLeuGly 45
           ||||||||| ||| |||||::::: |:: ||||||| ::: |||||
Db   209 CGGGTGATTAGGTACGTCGTTCGGATGTGGTTACGATCGTTTCGTCGTCGTTTAGGT 268

RESULT 14
ABQ46477/c
ID   ABQ46477 standard; DNA; 885 BP.
XX
AC   ABQ46477;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 33068.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.

XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 416 BP; 115 A; 74 C; 118 G; 109 T; 0 other;

Alignment Scores:
Pred. No.:   3.35   Length: 416
Score: 58.50   Matches: 18
Percent Similarity: 60.53%   Conservative: 5
Best Local Similarity: 47.37%   Mismatches: 13
Query Match: 25.77%   Indels: 2
                          24   Gaps: 1

US-09-833-017B-2 (1-46) x ABK77550 (1-416)

Qy   3 LysThrLeuSerLeuLysAsnAspPhe-LysGluIleLysThrAspGluLeuGluIleIl 22
           :::||| ||| |||::: :::|||:||:::
Db   36 AGAACATGAAGTGCTAAAGGGGATTACGACGGACATTCGCGAAAAGGAAGTTGTTGCAT 95

Qy   22 eIleGlyGlySerGlySer---LeuSerThrPhePheArgLeuPheAsnArg 38
           |||||||| |||||||| |||||||
Db   96 TATTGGCCCGTCTGGCTCCGGGAAAAGTACGTTCTCCGCGTTGTTTAAACCGG 147

RESULT 13
ABQ46476
ID   ABQ46476 standard; DNA; 885 BP.
XX
AC   ABQ46476;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 33067.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX

```
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated  The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 885 BP; 332 A; 327 C; 99 G; 127 T; 0 other;

Alignment Scores:
Pred. No.:              9.23        Length:        885
Score:                 58.50        Matches:        15
Percent Similarity:    41.67%       Conservative:   16
Best Local Similarity: 25.00%       Mismatches:     16
Query Match:           25.77%       Indels:         19
DB:                       24        Gaps:            1

US-09-833-017B-2 (1-46) x ABQ46477 (1-885)

Qy    5 LeuSerLeuLysAsnAspPheLysGluThrAspGluLeuGluIle--------  21
               ::|    ||||||:::::::::
Db  797 TTGGCGTTTCGTTTTCGTTTCGGAGGTTCGGGCGGGAGGTTGGGAGGTTGG 738

Qy   22 ------------------------------------IleIleGlyGly 25
                                               :: |||
Db  737 AGGCGAGTTCGACGCGAGAGATTGGGCGGATTGGTTGGATTGGTTGGACGGTTGGCGGGA 678

Qy   26 SerGlyLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 45
           |||  ||| ||| |||||||::::::   ||||||||
Db  677 CGGGGTGATTAGGTACGTCGTTCGCGATGTGGTTACGACGCGTTTCGTCGTTCGTTTAGGT 618

RESULT 15
ABN70018/c
ID   ABN70018 standard; DNA; 1530 BP.
XX
AC   ABN70018;
XX
DT   01-JUL-2002  (first entry)
XX
```

```
DE   Streptococcus polynucleotide SEQ ID NO 7949.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP29387.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3923-3924; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 1530 BP; 438 A; 313 C; 341 G; 438 T; 0 other;

Alignment Scores:
Pred. No.:              19.2        Length:        1530
Score:                 58.50        Matches:         16
Percent Similarity:    52.27%       Conservative:     7
Best Local Similarity: 36.36%       Mismatches:      20
Query Match:           25.77%       Indels:           1
DB:                       24        Gaps:             1

US-09-833-017B-2 (1-46) x ABN70018 (1-1530)

Qy    3 LysThrLeuSerLeuLysAsnAspPheLysGluLysThrAspGluLeuGluIleIle 22
            |||  |||  |||  ::||||||  |||     :::::|||
Db  993 AAAGTCTTGGACATCAAGAATTTTCACCAGTTCTTCTTCGAGCAACTCCACCATAGC 934

Qy   23 IleGlyGly--SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThr 41
            ||:|||    ||||||
Db  933 ATTGGTAATACCACCAAGAGCAAAGCTAGCTTTGATATTTCTTTAATCATAGCTTCACG 874

Qy   42 GlnAlaLeuGly 45
            |||
```

Wed Nov 13 10:42:11 2002

us-09-833-017b-2.rng

Db   873 CATGAAACGGGT 862

Search completed: November 8, 2002, 22:10:56
Job time : 233.313 secs

```
                    GenCore version 5.1.3
           Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:    November  8, 2002, 22:05:37 ; Search time 48.0597 Seconds
                                          (without alignments)
                                          293.533 Million cell updates/sec

Title:        US-09-833-017B-2
Perfect score: 227
Sequence:     1 MKKTLSLKNDFKEIKTDELE.........GSLSTPFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext  0.5
               Ygapop 10.0 , Ygapext  0.5
               Fgapop  6.0 , Fgapext  7.0
               Delop   6.0 , Delext   7.0

Searched:     441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:     882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105349_4858/app_query.fasta_1.398
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017 @CGN_1_1_17_@runat_05112002_105349_4858 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
         1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
         2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
         3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
         4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
         5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
         6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

                             %
Result          Query
No.     Score   Match   Length   DB  ID                   Description
-----   -----   -----   ------   --  ------------------   -----------
c  1    69      30.4    8411     4   US-08-961-527-16     Sequence 16, Appl
c  2    59      26.0    1481     4   US-09-434-774-1      Sequence 1, Appli
   3    55.5    24.4    1662     2   US-08-663-566A-12    Sequence 12, Appl
   4    55.5    24.4    1662     2   US-08-023-610-12     Sequence 12, Appl
   5    55.5    24.4    1662     2   US-08-288-065A-12    Sequence 12, Appl
   6    55.5    24.4    1662     4   US-08-362-240A-12    Sequence 12, Appl
   7    55.5    24.4    1662     4   US-08-804-372A-10    Sequence 10, Appl
   8    55.5    24.4    3825     5   PCT-US95-10245-12    Sequence 12, Appl
   9    55.5    24.4    5310678  5   5310678-2            Patent No. 5310678
   10   55.5    24.4    4177     2   US-08-484-575A-12    Sequence 12, Appl
   11   55.5    24.4    4177     3   US-08-477-459-12     Sequence 12, Appl
   12   55.5    24.4    4177     3   US-08-479-869-12     Sequence 12, Appl
   13   55.5    24.4    4177     3   US-08-486-414-12     Sequence 12, Appl
   14   55.5    24.4    4177     5   PCT-US94-01826A-12   Sequence 12, Appl
   15   55.5    24.4    4177     5   PCT-US94-02252A-12   Sequence 12, Appl
   16   55      24.2    3033     4   US-09-134-001C-2341  Sequence 2341, Ap
   17   54.5    24.0    5848     4   US-09-368-588-3      Sequence 3, Appli
c  18   54      23.8    5738     1   US-09-221-017B-246   Sequence 246, App
   19   54      23.8    5738     3   US-08-409-995-3      Sequence 3, Appli
   20   54      23.8    5738     3   US-08-685-467-3      Sequence 3, Appli
   21   54      23.8    7253     4   US-09-268-347-35     Sequence 35, Appl
   22   54      23.8    7291     4   US-08-913-942-3      Sequence 3, Appli
   23   54      23.8    29793    4   US-09-302-812-38     Sequence 38, Appl
   24   54      23.8    29793    4   US-09-511-477-38     Sequence 38, Appl
   25   54      23.6    1197     4   US-09-511-507-38     Sequence 38, Appl
c  26   53.5    23.6    1197     4   US-08-745-995A-10    Sequence 10, Appl
   27   53.5    23.6    1197     4   US-08-745-995A-11    Sequence 11, Appl
c  28   53.5    23.6    1358     4   US-08-745-995A-7     Sequence 7, Appli
   29   53.5    23.6    1358     4   US-08-745-995A-9     Sequence 9, Appli
c  30   53.5    23.6    1845     3   US-08-989-251-24     Sequence 24, Appl
   31   53.5    23.6    1845     3   US-09-940-350-24     Sequence 24, Appl
c  32   53.5    23.6    1845     4   US-09-528-108-24     Sequence 24, Appl
c  33   53.5    23.6    2023     3   US-08-989-251-36     Sequence 36, Appl
c  34   53.5    23.6    2023     3   US-09-340-250-36     Sequence 36, Appl
c  35   53.5    23.6    2023     4   US-09-528-108-36     Sequence 36, Appl
c  36   53      23.3    1260     2   US-08-578-158-1      Sequence 1, Appli
   37   53      23.3    10523    4   US-09-453-702B-13    Sequence 13, Appl
   38   53      23.3    12720    1   US-08-403-866-11     Sequence 11, Appl
c  39   52      22.9    861      1   US-07-877-516-2      Sequence 2, Appli
   40   52      22.9    1242     1   US-07-877-516-1      Sequence 1, Appli
c  41   52      22.9    1947     3   US-08-604-991-1      Sequence 1, Appli
c  42   52      22.9    1947     3   US-09-363-639-1      Sequence 1, Appli
c  43   52      22.9    8959     1   US-08-920-812-1      Sequence 1, Appli
c  44   52      22.9    8959     1   US-08-920-827-1      Sequence 1, Appli
c  45   52      22.9    8959     1   US-08-921-177-1      Sequence 1, Appli

                              ALIGNMENTS

RESULT 1
US-08-961-527-16/C
; Sequence 16, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
;  APPLICANT: Charles Kunsch
;  TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;  NUMBER OF SEQUENCES: 391
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/961,527
;   FILING DATE:
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P1
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
;  INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
```

```
        LENGTH: 8411 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
US-08-961-527-16

Alignment Scores:
Pred. No.:              0.752
Score:                  69.00
Percent Similarity:     50.88%
Best Local Similarity:  33.33%
Query Match:            30.40%
DB:                     4

US-09-833-017B-2 (1-46) x US-08-961-527-16 (1-8411)

Qy    2 LysLysThrLeuSerLeuLysAsnAspPhe------------------------LysGlu 13
          |||||||   ::   |||                                     |||||
Db 3566 AAAAGACTTTTACAACTGATAATGTCTTCCTTCTTAATCTAGGATATACTAAAGAG 3507

Qy   14 IleLeuThrAspGluLeuGluIleIleIleGlyGlySerGlySerLeuSer--------- 30
          |||||                         |||:|||           ::|||
Db 3506 AAAAAAACTTTCAATATCAATGTTAGTTTCCCTGGAAATTCACAAGTATCATTTGAATCT 3447

Qy   31 ---ThrPhePheArgLeuAsnArgSerPheThrGlnAlaLeuGlyLys 46
             ||||||::|||||          ::::|||||||||||:::   |||
Db 3446 CCTACCTTCTATCGTTTAGATACCAAAACTTTCACCGAGGCAATTCAAAAA 3396


RESULT 2
US-09-434-774-1/c
; Sequence 1, Application US/09434774A
; Patent No. 6235514
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL
; TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE
; FILE REFERENCE: wsurl4448
; CURRENT APPLICATION NUMBER: US/09/434,774A
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1481
; TYPE: DNA
; ORGANISM: Mentha piperita
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(1217)
US-09-434-774-1

Alignment Scores:
Pred. No.:              3.25
Score:                  59.00
Percent Similarity:     56.10%
Best Local Similarity:  41.46%
Query Match:            25.99%
DB:                     4

US-09-833-017B-2 (1-46) x US-09-434-774-1 (1-1481)

Qy    1 MetLysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
          :::||||||||||||   ::   ::            |||         ||||   |||
Db  334 CTTAAGAAAACATTAATCTTGCAAGGCGAA--------------AAAAGAGTGAGTCTTGAA 287

Qy   21 IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuAsnArgSerPhe 40
          |||   |||   :::  |||   ||| |||::|||||||||:::||||||||:::
Db  286 ATCCCAGCCTCCCTATCCACTTCATCAGCTAATTTGTTTAATTTATTCCAAGATCATAT 227

Qy   41 Thr 41
          |||
Db  226 ACC 224


RESULT 3
US-08-663-566A-12
; Sequence 12, Application US/0866566A
; Patent No. 5853733
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Macdonald, Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
; TITLE OF INVENTION: and Uses Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,566A
; FILING DATE: June 13, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1662 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1662
US-08-663-566A-12

Alignment Scores:
Pred. No.:              13.8
Score:                  55.50
Percent Similarity:     47.92%
Best Local Similarity:  33.33%
Query Match:            24.45%
DB:                     2

US-09-833-017B-2 (1-46) x US-08-663-566A-12 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe--------------LysGluIleLysThrAspGlu 18
          ||||||   |||   :::||                  ::||| |||   |||
Db 1297 ACTTTAAGGCTCAGTGGGGAATTCGATGTAACTTATCAGAAGAAATATCCAATACAAGAT 1356

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuAsnArg 38
          ::|||||| |||   ::|||||| |||   |||||  :||:
Db 1357 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
          ||| ||| :::   ||||||||::
Db 1417 TCGATCAGTAATGCCTTGAATAAG 1440


RESULT 4
US-08-023-610-12
; Sequence 12, Application US/08023610
; Patent No. 5928648
```

```
GENERAL INFORMATION:
     APPLICANT: Cochran Ph.D, Mark D
     APPLICANT: Macdonald Ph.D., Richard D
     TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
     TITLE OF INVENTION: and Uses Thereof
     NUMBER OF SEQUENCES: 47
     CORRESPONDENCE ADDRESS:
        ADDRESSEE: John P. White
        STREET: 30 Rockefeller Plaza
        CITY: New York
        STATE: New York
        COUNTRY: USA
        ZIP: 10112
     COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/023,610
        FILING DATE: February 26, 1993
        CLASSIFICATION: 435
     ATTORNEY/AGENT INFORMATION:
        NAME: White Esq, John P
     TELECOMMUNICATION INFORMATION:
        TELEPHONE: (212)977-9550
        TELEFAX: (212)664-0525
        TELEX: 422523
     INFORMATION FOR SEQ ID NO: 12:
     SEQUENCE CHARACTERISTICS:
        LENGTH: 1662 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
     MOLECULE TYPE: DNA (genomic)
     HYPOTHETICAL: NO
     ANTI-SENSE: NO
     FEATURE:
        NAME/KEY: CDS
        LOCATION: 1..1662
US-08-023-610-12

Alignment Scores:
Pred. No.:                               Length:        1662
Score:                 13.8              Matches:       16
Percent Similarity:    55.50             Conservative:  7
Best Local Similarity: 47.92%            Mismatches:    20
Query Match:           33.33%            Indels:        5
DB:                    24.45%            Gaps:          1

US-09-833-017B-2 (1-46) x US-08-023-610-12 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
        ||||||  |||               ||  :||
Db 1297 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 1356

Qy   19 LeuGluIleIleIleGlyGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
        :::::|||||:|||||  |||        :::||||||
Db 1357 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        ||| ::: ||||||  |||
Db 1417 TCGATCAGTAATGCCTTGAATAAG 1440

RESULT 5
US-08-288-065A-12
; Sequence 12, Application US/08288065A
; Patent No. 5961982
  GENERAL INFORMATION:
     APPLICANT: Cochran, Mark D
     APPLICANT: Macdonald, Richard D
     TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-
     TITLE OF INVENTION: HVT-050 and Uses Thereof
```

```
     NUMBER OF SEQUENCES: 56
     CORRESPONDENCE ADDRESS:
        ADDRESSEE: John P. White
        STREET: 1185 Avenue of the Americas
        CITY: New York
        STATE: New York
        COUNTRY: USA
        ZIP: 10036
     COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/288,065A
        FILING DATE: Aug-09-94
        CLASSIFICATION: 435
     ATTORNEY/AGENT INFORMATION:
        NAME: White, John P
     TELECOMMUNICATION INFORMATION:
        TELEPHONE: (212)278-0400
        TELEFAX: (212)391-0526
        TELEX: 422523
     INFORMATION FOR SEQ ID NO: 12:
     SEQUENCE CHARACTERISTICS:
        LENGTH: 1662 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
     MOLECULE TYPE: DNA (genomic)
     HYPOTHETICAL: NO
     ANTI-SENSE: NO
     FEATURE:
        NAME/KEY: CDS
        LOCATION: 1..1662
US-08-288-065A-12

Alignment Scores:
Pred. No.:                               Length:        1662
Score:                 13.8              Matches:       16
Percent Similarity:    55.50             Conservative:  7
Best Local Similarity: 47.92%            Mismatches:    20
Query Match:           33.33%            Indels:        5
DB:                    24.45%            Gaps:          1

US-09-833-017B-2 (1-46) x US-08-288-065A-12 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
        ||||||  |||               ||  :||
Db 1297 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 1356

Qy   19 LeuGluIleIleIleGlyGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
        :::::|||||:|||||  |||        :::||||||
Db 1357 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        ||| ::: ||||||  |||
Db 1417 TCGATCAGTAATGCCTTGAATAAG 1440

RESULT 6
US-08-362-240A-12
; Sequence 12, Application US/08362240A
; Patent No. 5965138
  GENERAL INFORMATION:
     APPLICANT: Cochran, Mark D
     APPLICANT: Junker, David
     APPLICANT: Wild, Martha A
     TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
     NUMBER OF SEQUENCES: 60
     CORRESPONDENCE ADDRESS:
        ADDRESSEE: John P. White
        STREET: 1185 Avenue of the Americas
        CITY: New York
```

```
                STATE: New York
                COUNTRY: USA
                ZIP: 10036
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: IBM PC compatible
                OPERATING SYSTEM: PC-DOS/MS-DOS
                SOFTWARE: PatentIn Release #1.0, Version #1.25
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: US/08/362,240A
                FILING DATE: Dec-22-94
                CLASSIFICATION: 435
        ATTORNEY/AGENT INFORMATION:
                NAME: White, John P
                REGISTRATION NUMBER: 28,678
        TELECOMMUNICATION INFORMATION:
                TELEPHONE: (212)278-0400
                TELEFAX: (212)391-0526
                TELEX: 422523
INFORMATION FOR SEQ ID NO: 12:
        SEQUENCE CHARACTERISTICS:
                LENGTH: 1662 base pairs
                TYPE: nucleic acid
                STRANDEDNESS: double
                TOPOLOGY: linear
        MOLECULE TYPE: DNA (genomic)
        HYPOTHETICAL: NO
        ANTI-SENSE: NO
        FEATURE:
                NAME/KEY: CDS
                LOCATION: 1..1662
US-08-362-240A-12

Alignment Scores:
Pred. No.:                13.8        Length:        1662
Score:                    55.50      Matches:       16
Percent Similarity:       47.92%     Conservative:  7
Best Local Similarity:    33.33%     Mismatches:    20
Query Match:              24.45%     Indels:        5
DB:                                  Gaps:          2
US-08-362-240A-12

US-09-833-017B-2 (1-46) x US-08-362-240A-12 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
        ||||||  ||       ||         ||||  :::
Db 1297 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 1356

Qy   19 LeuGluIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
        :::::||||| |||:::  |||:::  ::||||||
Db 1357 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        ||| :::   ||||||| |||
Db 1417 TCGATCAGTAATGCCTTGAATAAG 1440

RESULT 7
US-08-804-372A-10
; Sequence 10, Application US/08804372A
; Patent No. 6183753
; GENERAL INFORMATION:
        APPLICANT: Cochran, Mark D.
        APPLICANT: Wild, Martha A.
        APPLICANT: Winslow, Barbara J.
        TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses
        TITLE OF INVENTION: Thereof
        NUMBER OF SEQUENCES: 32
        CORRESPONDENCE ADDRESS:
                ADDRESSEE: Cooper & Dunham LLP
                STREET: 1185 Avenue of the Americas
                CITY: New York
                STATE: New York
                COUNTRY: USA
                ZIP: 10036
```

```
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: IBM PC compatible
                OPERATING SYSTEM: PC-DOS/MS-DOS
                SOFTWARE: PatentIn Release #1.0, Version #1.30
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: US/08/804,372A
                FILING DATE:
                CLASSIFICATION: 424
        ATTORNEY/AGENT INFORMATION:
                NAME: White, John P.
                REGISTRATION NUMBER: 28,678
                REFERENCE/DOCKET NUMBER: 2552/39115E
        TELECOMMUNICATION INFORMATION:
                TELEPHONE: (212) 278-0400
                TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 10:
        SEQUENCE CHARACTERISTICS:
                LENGTH: 1662 base pairs
                TYPE: nucleic acid
                STRANDEDNESS: double
                TOPOLOGY: linear
        MOLECULE TYPE: DNA (genomic)
        HYPOTHETICAL: NO
        ANTI-SENSE: NO
        FEATURE:
                NAME/KEY: CDS
                LOCATION: 1..1662
US-08-804-372A-10

Alignment Scores:
Pred. No.:                13.8        Length:        1662
Score:                    55.50      Matches:       16
Percent Similarity:       47.92%     Conservative:  7
Best Local Similarity:    33.33%     Mismatches:    20
Query Match:              24.45%     Indels:        5
DB:                                  Gaps:          1
US-08-804-372A-10

US-09-833-017B-2 (1-46) x US-08-804-372A-10 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
        ||||||  ||       ||         ||||  :::
Db 1297 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 1356

Qy   19 LeuGluIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
        :::::||||| |||:::  |||:::  ::||||||
Db 1357 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        ||| :::   ||||||| |||
Db 1417 TCGATCAGTAATGCCTTGAATAAG 1440

RESULT 8
PCT-US95-10245-12
; Sequence 12, Application PCT/TUS9510245
; GENERAL INFORMATION:
        APPLICANT: SYNTRO CORPORATION
        TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof
        NUMBER OF SEQUENCES: 60
        CORRESPONDENCE ADDRESS:
                ADDRESSEE: John P. White
                STREET: 1185 Avenue of the Americas
                CITY: New York
                STATE: New York
                COUNTRY: USA
                ZIP: 10036
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: IBM PC compatible
                OPERATING SYSTEM: PC-DOS/MS-DOS
                SOFTWARE: PatentIn Release #1.0, Version #1.25
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: PCT/US95/10245
```

```
        FILING DATE: 09-AUG-1995
      CLASSIFICATION:
      ATTORNEY/AGENT INFORMATION:
        NAME: White, John P
        REGISTRATION NUMBER: 28,678
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (212)278-0400
        TELEFAX: (212)391-0526
        TELEX: 422523
      INFORMATION FOR SEQ ID NO: 12:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 1662 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
        MOLECULE TYPE: DNA (genomic)
        ANTI-SENSE: NO
        HYPOTHETICAL: NO
        FEATURE:
        NAME/KEY: CDS
        LOCATION: 1..1662
      PCT-US95-10245-12

Alignment Scores:
Pred. No.:                13.8        Length:        1662
Score:                    55.50      Matches:         16
Percent Similarity:       47.92%     Conservative:     7
Best Local Similarity:    33.33%     Mismatches:      20
Query Match:              24.45%     Indels:           5
DB:                       5          Gaps:             1

US-09-833-017B-2 (1-46) x PCT-US95-10245-12 (1-1662)

Qy    4 ThrLeuSerLeuLysAsnAspPhe----------LysGluIleLysThrAspGlu 18
        ||||||  |||                         |||  |||      :::|||
Db 1297 ACTTTAAGGCTCAGTGGGAATTCGATGAACTTATCAGAAGAATATCTCAATACAAGAT 1356

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
        :::||||||| ||| ::|                 ::|||||
Db 1357 TCTCAAGTAATAATAACAGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 1416

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        |||  :::     |||||
Db 1417 TCGATCAGTAATGCTTTGAATAAG 1440

RESULT 9
5310678-2
;Patent No. 5310678
;T.;Millar, Neil S.
;  APPLICANT: Bingham, Richard W.;Chambers, Philip;Emmerson, Peter
;  TITLE OF INVENTION: NEWCASTLE DISEASE VIRUS GENE CLONES
;  NUMBER OF SEQUENCES: 3
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/07/438,945
;  FILING DATE: 17-NOV-1989
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 885,765
;  FILING DATE: 15-JUL-1986
;SEQ ID NO:2:
; LENGTH: 3825
5310678-2

Alignment Scores:
Pred. No.:                40.2        Length:        3825
Score:                    55.50      Matches:         16
Percent Similarity:       47.92%     Conservative:     7
Best Local Similarity:    33.33%     Mismatches:      20
Query Match:              24.45%     Indels:           5
DB:                       6          Gaps:             1

US-09-833-017B-2 (1-46) x 5310678-2 (1-3825)

Qy    4 ThrLeuSerLeuLysAsnAspPhe----------LysGluIleLysThrAspGlu 18
```

```
        ||||||  |||                         :::|||
Db 1343 ACTTTAAGGCTCAGTGGGAATTCGATGCAACTTATCAGAAGAATATCTCAATACAAGAT 1402

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
        :::||||||| ||| ::|                 ::|||||
Db 1403 TCTCAAGTAATAATAACAGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 1462

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
        |||  :::     |||||
Db 1463 TCGATCAGTAATGCTTTGAATAAG 1486

RESULT 10
US-08-484-575A-12
; Sequence 12, Application US/08484575A
; Patent No. 5925358
;  GENERAL INFORMATION:
;  APPLICANT: Mark D. Cochran and David E. Junker
;  TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
;  NUMBER OF SEQUENCES: 42
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: John P. White
;  STREET: 1185 Avenue of the Americas
;  CITY: New York
;  STATE: New York
;  COUNTRY: USA
;  ZIP: 10036
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/484,575A
;  FILING DATE: 07-JUN-1995
;  CLASSIFICATION: 424
;  ATTORNEY/AGENT INFORMATION:
;  NAME: White Esq, John P
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (212)278-0450
;  TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 12:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 4177 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: double
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: 115..1860
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: 2095..3756
US-08-484-575A-12

Alignment Scores:
Pred. No.:                45          Length:        4177
Score:                    55.50       Matches:         16
Percent Similarity:       47.92%      Conservative:     7
Best Local Similarity:    33.33%      Mismatches:      20
Query Match:              24.45%      Indels:           5
DB:                       2           Gaps:             1

US-09-833-017B-2 (1-46) x US-08-484-575A-12 (1-4177)

Qy    4 ThrLeuSerLeuLysAsnAspPhe----------LysGluIleLysThrAspGlu 18
        ||||||  |||                         |||  |||      :::|||
Db 3391 ACTTTAAGGCTCAGTGGGAATTCGATGAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
        :::||||||| ||| ::|                 ::|||||
```

Db 3451 TCTCAAGTAATAATAAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy   39 SerPheThrGlnAlaLeuGlyLys 46

Db 3511 TCGATCAGTAATGCCTTGAATAAG 3534

RESULT 11
US-08-477-459-12
; Sequence 12, Application US/08477459
; Patent NO. 6001369
; GENERAL INFORMATION:
;   APPLICANT: Mark D. Cochran
;   TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: John P. White
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/477,459
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White Esq, John P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 278-0400
;   TELEFAX: (212) 391-0525
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4177 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 115..1860
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2095..3756
US-08-477-459-12

Alignment Scores:
Pred. No.:   45
Score: 55.50
Percent Similarity: 47.92%
Best Local Similarity: 33.33%
Query Match: 24.45%
DB: 3

Length: 4177
Matches: 16
Conservative: 7
Mismatches: 20
Indels: 5
Gaps: 1

US-09-833-017B-2 (1-46) x US-08-477-459-12 (1-4177)

Qy   4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
          ||||| ||| ||:|||| ||| ::||
Db 3391 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
          ::::||||||| |||::::   :::|||||||
Db 3451 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
          ||| ::: |||||| |||

Db 3511 TCGATCAGTAATGCCTTGAATAAG 3534

RESULT 12
US-08-479-869-12
; Sequence 12, Application US/08479869
; Patent No. 6123949
; GENERAL INFORMATION:
;   APPLICANT: Cochran Ph.D, Mark D
;   TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and
;   TITLE OF INVENTION: Uses Thereof
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: John P. White
;   STREET: 30 Rockefeller Plaza
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10112
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/479,869
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/024,156
;   FILING DATE: 26-FEB-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White Esq, John P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212)977-9550
;   TELEFAX: (212)664-0525
;   TELEX: 422523
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4177 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 115..1860
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2095..3756
US-08-479-869-12

Alignment Scores:
Pred. No.:   45
Score: 55.50
Percent Similarity: 47.92%
Best Local Similarity: 33.33%
Query Match: 24.45%
DB: 3

Length: 4177
Matches: 16
Conservative: 7
Mismatches: 20
Indels: 5
Gaps: 1

US-09-833-017B-2 (1-46) x US-08-479-869-12 (1-4177)

Qy   4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
          ||||| ||| ||:|||| ||| ::||
Db 3391 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy   19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
          ::::||||||| |||::::   :::|||||||
Db 3451 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy   39 SerPheThrGlnAlaLeuGlyLys 46
          ||| ::: |||||| |||

Db 3511 TCGATCAGTAATGCCTTGAATAAG 3534

RESULT 13
US-08-486-414-12
; Sequence 12, Application US/08486414B
; Patent No. 6135318
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
; FILE REFERENCE: 42771D
; CURRENT APPLICATION NUMBER: US/08/486,414B
; CURRENT FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 12
; LENGTH: 4177
; TYPE: DNA
; ORGANISM: Fowlpox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)..(1857)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2095)..(3753)
US-08-486-414-12

Alignment Scores:
Pred. No.: 45
Score: 55.50
Percent Similarity: 47.92%
Best Local Similarity: 33.33%
Query Match: 24.45%
DB: 3

Length: 4177
Matches: 16
Conservative: 7
Mismatches: 20
Indels: 5
Gaps: 1

US-09-833-017B-2 (1-46) x US-08-486-414-12 (1-4177)

Qy 4 ThrLeuSerLeuLysAsnAspPhe--------------LysGluIleLysThrAspGlu 18
Db 3391 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy 19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
Db 3451 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy 39 SerPheThrGlnAlaLeuGlyLys 46
Db 3511 TCGATCAGTAATGCCTTGAATAAG 3534

RESULT 14
PCT-US94-01826A-12
; Sequence 12, Application PCT/US9401826A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01826A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..1860
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2095..3756
PCT-US94-01826A-12

Alignment Scores:
Pred. No.: 45
Score: 55.50
Percent Similarity: 47.92%
Best Local Similarity: 33.33%
Query Match: 24.45%
DB: 5

Length: 4177
Matches: 16
Conservative: 7
Mismatches: 20
Indels: 5
Gaps: 1

US-09-833-017B-2 (1-46) x PCT-US94-01826A-12 (1-4177)

Qy 4 ThrLeuSerLeuLysAsnAspPhe--------------LysGluIleLysThrAspGlu 18
Db 3391 ACTTTAAGGCTCAGTGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy 19 LeuGluIleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
Db 3451 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy 39 SerPheThrGlnAlaLeuGlyLys 46
Db 3511 TCGATCAGTAATGCCTTGAATAAG 3534

RESULT 15
PCT-US94-02252A-12
; Sequence 12, Application PC/TUS9402252A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02252A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523

```
;  INFORMATION FOR SEQ ID NO:    12:
;  SEQUENCE CHARACTERISTICS:
;     LENGTH:  4177 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;  HYPOTHETICAL:  NO
;  ANTI-SENSE:  NO
;  FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  115..1860
;  FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  2095..3756
PCT-US94-02252A-12
```

```
Alignment Scores:
Pred. No.:                45                  Length:        4177
Score:                    55.50               Matches:       16
Percent Similarity:       47.92%              Conservative:  7
Best Local Similarity:    33.33%              Mismatches:    20
Query Match:              24.45%              Indels:        5
DB:                       5                   Gaps:          1
```

```
US-09-833-017B-2 (1-46) x PCT-US94-02252A-12 (1-4177)

Qy      4 ThrLeuSerLeuLysAsnAspPhe--------------LysGluIleLysThrAspGlu 18
            ||||||  |||      ::|||            ||| |||        :::
Db   3391 ACTTTAAGGCTCAGTGGGGAATTCGATGTAACTTATCAGAAGAATATCTCAATACAAGAT 3450

Qy     19 LeuGluIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArg 38
            :::::||||||| |||:::       ::|||||
Db   3451 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 3510

Qy     39 SerPheThrGlnAlaLeuGlyLys 46
            ||| :::     ||||||  |||
Db   3511 TCGATCAGTAATGCCTTGAATAAG 3534
```

```
Search completed: November  8, 2002, 23:41:15
Job time : 53.0597 secs
```

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2002, 22:57:42 ; Search time 45.3134 Seconds
                                        (without alignments)
                                        360.161 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 227
Sequence: 227
             1 MKKTLSLKNDFKEIKTDELE........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop   6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105351_4940/runat_05112002_105351_4940/app_query.fasta_1.398
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09833017_@CGN_1_18_@runat_05112002_105351_4940
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA:*
           1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
           2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
           3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
           4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
           5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
           6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
           7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
           8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
           9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
          10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
          11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
          12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
          13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
          14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	227	100.0	141	10	US-09-833-017-1	Sequence 1, Appli
2	227	100.0	2557	10	US-09-833-017-21	Sequence 21, Appl
3	58.5	25.8	416	10	US-09-974-300-4841	Sequence 4841, Ap
c 4	57	25.1	942	10	US-09-974-300-2095	Sequence 2095, Ap

c	5	57	25.1	48667	10	US-09-822-268A-3	Sequence 3, Appli
	6	56.5	24.9	843	10	US-09-974-300-3171	Sequence 3171, Ap
	7	56.5	24.9	13884	10	US-09-070-927A-341	Sequence 341, App
	8	56	24.7	2820	10	US-09-815-242-6700	Sequence 6700, Ap
c	9	56	24.7	5480	10	US-09-070-927A-132	Sequence 132, App
	10	56	24.7	465237	10	US-09-931-267A-1	Sequence 1, Appli
	11	55.5	24.4	3570	10	US-09-881-457A-1	Sequence 1, Appli
c	12	55	24.2	3218	10	US-09-788-657-5	Sequence 5, Appli
	13	54.5	24.0	11871	10	US-09-070-927A-171	Sequence 171, App
	14	54	23.8	2376	10	US-09-917-800A-1496	Sequence 1496, Ap
	15	54	23.8	29793	10	US-09-973-451-38	Sequence 38, Appl.
c	16	53.5	23.6	1845	10	US-09-921-398-24	Sequence 24, Appl
c	17	53.5	23.6	2023	10	US-09-921-398-36	Sequence 36, Appl
c	18	53	23.3	438	10	US-09-924-035A-605	Sequence 605, App
	19	53	23.3	971	10	US-09-770-445-285	Sequence 285, App
	20	52.5	23.1	24601	10	US-09-070-927A-223	Sequence 223, App
	21	52	22.9	269	10	US-09-294-093B-1839	Sequence 1839, Ap
	22	52	22.9	275	10	US-09-878-574-15440	Sequence 15440, A
c	23	52	22.9	10476	10	US-09-964-824A-98	Sequence 98, Appl
c	24	52	22.9	10476	10	US-09-964-824A-552	Sequence 552, App
	25	51.5	22.7	395	10	US-09-960-352-12338	Sequence 12338, A
	26	51.5	22.7	421	10	US-09-880-107-1416	Sequence 1416, Ap
	27	51.5	22.7	774	10	US-09-741-669-184	Sequence 184, App
	28	51.5	22.7	921	10	US-09-912-020-115	Sequence 115, App
	29	51.5	22.7	1827	10	US-09-887-576-817	Sequence 817, App
	30	51.5	22.7	2979	9	US-09-938-842A-48	Sequence 48, Appl
c	31	51.5	22.7	7237	10	US-09-070-927A-239	Sequence 239, App
	32	51.5	22.7	16870	10	US-09-070-927A-98	Sequence 98, Appl
	33	51.5	22.7	147309	10	US-09-742-312-3	Sequence 3, Appli
	34	51	22.5	263	10	US-09-878-574-15097	Sequence 15097, A
c	35	51	22.5	492	10	US-09-783-590-1335	Sequence 1335, Ap
	36	51	22.5	663	10	US-09-815-242-8789	Sequence 8789, Ap
c	37	51	22.5	1726	9	US-09-938-842A-4596	Sequence 4596, Ap
	38	51	22.5	2979	10	US-09-815-242-4748	Sequence 4748, Ap
c	39	51	22.5	12822	10	US-09-764-847-1579	Sequence 1579, Ap
	40	51	22.5	32193	10	US-09-764-877-2623	Sequence 2623, Ap
	41	50.5	22.2	333	10	US-09-815-242-4774	Sequence 4774, Ap
	42	50.5	22.2	360	10	US-09-815-242-8884	Sequence 8884, Ap
	43	50.5	22.2	360	10	US-09-815-242-9051	Sequence 9051, Ap
	44	50.5	22.2	408	10	US-09-878-574-3898	Sequence 3898, Ap
c	45	50.5	22.2	425	10	US-09-815-242-1718	Sequence 1718, Ap

ALIGNMENTS

RESULT 1
US-09-833-017-1
; Sequence 1, Application US/09833017
; Patent No. US2002008130 2A1
; GENERAL INFORMATION:
; APPLICANT: CVITKOVITCH, Dennis
; TITLE OF INVENTION: SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATME
; TITLE OF INVENTION: OF CARIES
; FILE REFERENCE: P04885US1
; CURRENT APPLICATION NUMBER: US/09/833,017
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/CA00/00605
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 2,302,861
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 2,332,733
; PRIOR FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 1
;   LENGTH: 141
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(141)
US-09-833-017-1

```
Alignment Scores:
Pred. No.:              1.09e-29            Length:        141
Score:                  227.00             Matches:       46
Percent Similarity:     100.00%            Conservative:  0
Best Local Similarity:  100.00%            Mismatches:    0
Query Match:            100.00%            Indels:        0
DB:                                        Gaps:          10

US-09-833-017B-2 (1-46) x US-09-833-017-1 (1-141)

Qy   1 MetLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 60

Qy  21 IleIleIleGlyGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTT 120

Qy  41 ThrGlnAlaLeuGlyLys 46
        |||||||||||||||||
Db 121 ACACAAGCTTTGGGAAAA 138
```

RESULT 2
US-09-833-017-21
; Sequence 21, Application US/09833017
; Patent No. US2002008302A1
; GENERAL INFORMATION:
; APPLICANT: CVITKOVITCH, Dennis
; TITLE OF INVENTION: SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF CARIES
; FILE REFERENCE: P04885US1
; CURRENT APPLICATION NUMBER: US/09/833,017
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/CA00/00605
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 2,302,861
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 2,332,733
; PRIOR FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2557
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2557)
US-09-833-017-21

```
Alignment Scores:
Pred. No.:              5.95e-28            Length:        2557
Score:                  227.00             Matches:       46
Percent Similarity:     100.00%            Conservative:  0
Best Local Similarity:  100.00%            Mismatches:    0
Query Match:            100.00%            Indels:        0
DB:                                        Gaps:          10

US-09-833-017B-2 (1-46) x US-09-833-017-21 (1-2557)

Qy   1 MetLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAGAG 160

Qy  21 IleIleIleGlyGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161 ATTATCATTGGCGGAAGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTT 220

Qy  41 ThrGlnAlaLeuGlyLys 46
        |||||||||||||||||
Db 221 ACACAAGCTTTGGGAAAA 238
```

RESULT 3

US-09-974-300-4841
; Sequence 4841, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4841
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-4841

```
Alignment Scores:
Pred. No.:              0.438              Length:        416
Score:                  58.50             Matches:       18
Percent Similarity:     60.53%            Conservative:  5
Best Local Similarity:  47.37%            Mismatches:    13
Query Match:            25.77%            Indels:        2
DB:                                        Gaps:          1

US-09-833-017B-2 (1-46) x US-09-974-300-4841 (1-416)

Qy   3 LysThrLeuSerLeuLysAsnAspPhe-LysGluIleLysThrAspGluLeuGluIleIl 22
         ::||| ||| ||| ||   :::|| :::||:|||
Db  36 AGAACATGAAGTGCTAAAGGGATTACGACGGACATTCGCGAAAAGGAAGTTGTTTGCAT 95

Qy  22 eIleGlyGlySerGlySer--LeuSerThrPhePheArgLeuPheAsnArg 38
        ||||||||  ||||||||    |||||||
Db  96 TATTGGCCCGTCGGCTCCGGAAAAGTACGTTTCTCCCGTGTTTAAACCGG 147
```

RESULT 4
US-09-974-300-2095/c
; Sequence 2095, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 2095
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2095
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2095

```
Alignment Scores:
Pred. No.:              2.4                Length:        942
Score:                  57.00             Matches:       16
Percent Similarity:     53.49%            Conservative:  7
Best Local Similarity:  37.21%            Mismatches:    16
Query Match:            25.11%            Indels:        4
DB:                                        Gaps:          1
```

US-09-833-017B-2 (1-46) x US-09-974-300-2095 (1-942)

Qy   2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle 21
        ::::||| ||| ::|||::|||:::|||
Db 718 GAGAAAATGCTGCAGCATCGCCGTTTTAAAAGGTCCGCCGTACTGATGATTTGGTT--- 662

Qy   22 IleIleGlyGlySerGlyLeuSerThrPheArgLeuPheAsnArgSerPheThr 41
        |||||||||::||| ||||| ||||| ::: |||
Db 661 --------GGCTCCGGCGGCGCTCGGCATTTTTTTCTGCTTTTCCGTCAAGATTTCTAC 611

Qy   42 GlnAlaLeu 44
        |||:::
Db 610 AACGCAATA 602

RESULT 5
US-09-822-268A-3/c
; Sequence 3, Application US/09822268A
; Patent No. US20020048787A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000667
; CURRENT APPLICATION NUMBER: US/09/822,268A
; CURRENT FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/211,387
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 48667
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(48667)
; OTHER INFORMATION: n = A,T,C or G
US-09-822-268A-3

Alignment Scores:
Pred. No.:   555   Length: 48667
Score: 57.00   Matches: 12
Percent Similarity: 62.07%   Conservative: 6
Best Local Similarity: 41.38%   Mismatches: 11
Query Match: 25.11%   Indels: 0
DB: Gaps: 0

US-09-833-017B-2 (1-46) x US-09-822-268A-3 (1-48667)

Qy   16 ThrAspGluLeuGluIleIleIleGlyGlySerGlyLeuSerThrPhePheArgLeu 35
        ||| :::||| ||| |||||||||||||||
Db 15239 ACCCAGCTAATTTTCATACTTTTGGTAGAGATGGGAGTCTCTCTACGTTGCTCAAGCTG 15180

Qy   36 PheAsnArgSerPheThrGlnAlaLeu 44
        ::: ||||||| ||||||:::
Db 15179 ATCTCAAATTCCTGGGCTCAAGCGATC 15153

RESULT 6
US-09-974-300-3171
; Sequence 3171, Application US/09974300
; Patent No. US2002014672lA1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3171
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3171

Alignment Scores:
Pred. No.:   2.49   Length: 843
Score: 56.50   Matches: 16
Percent Similarity: 48.89%   Conservative: 6
Best Local Similarity: 35.56%   Mismatches: 14
Query Match: 24.89%   Indels: 9
DB: 10   Gaps: 1

US-09-833-017B-2 (1-46) x US-09-974-300-3171 (1-843)

Qy   2 LysLysThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIle 21
        ||| |||||| ::::||| |||||||:::   :::
Db 400 AAATTAACCCTATCATCGAGACGCTAAGAAACGAGTTGAAACTTGATGAGGTACCGTTG 459

Qy   22 IleIleGlyGlySerGlySerLeu------------------------SerThrPhe 32
        ||||||||||||| ||| :::|||
Db 460 ATTATTGGAGGGCTTGGTGACTTTCTTGGGAAGACCGGTTTTGGACGACCACCGACCGAA 519

Qy   33 PheArgLeuPheAsn 37
        |||||| |||
Db 520 TTTCGACAGGTTAAT 534

RESULT 7
US-09-070-927A-341
; Sequence 341, Application US/09070927A
; Patent No. US2002012011GA1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;   Patrick J. Dillon
;   Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/070,927A
;   FILING DATE: 04-May-2000
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/046,655
;   FILING DATE: 1997-05-16
;   APPLICATION NUMBER: 60/044,031
;   FILING DATE: 1997-05-06
;   APPLICATION NUMBER: 60/066,009
;   FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
;   NAME: Kenley K. Hoover
;   REGISTRATION NUMBER: 40,302
;   REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 341:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 13884 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 341:
US-09-070-927A-341

Alignment Scores:
Pred. No.:   119   Length: 13884
Score: 56.50   Matches: 18
Percent Similarity: 54.76%   Conservative: 5
Best Local Similarity: 42.86%   Mismatches: 14
Query Match: 24.89%   Indels: 5
DB: 10   Gaps: 2

US-09-833-017B-2 (1-46) x US-09-070-927A-341 (1-13884)

Qy   1 MetLysThrLeuSer--------LeuLysAsnAspPheLysGluIleLysThr 16
           :::::|||||| |||||:::
Db 1233 TTACAGAAAACATTCGGAAAAAATGAAGTACTTAAGGAATTGACTTAGAAGCT 1292

Qy   17 AspGluLeuGluIleIleIleGlyGlySerGlySer--LeuSerThrPheArgLeu 35
           ||||:::   :::||||||||| |||||||
Db 1293 GGTGAAGTGTGCGTAATTATTGGCCCTTCTGGTAGTGGTAAAAGTACTTTTTTACGTTGC 1352

Qy   36 PheAsn 37
           |||
Db 1353 TTGAAC 1358

RESULT 8
US-09-815-242-6700
; Sequence 6700, Application US/09815242
; Patent No. US2002006156A1
; GENERAL INFORMATION:
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari L.
;   APPLICANT: Zyskind, Judith W.
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John D.
;   APPLICANT: Carr, Grant J.
;   APPLICANT: Yamamoto, Robert T.
;   APPLICANT: Xu, H. Howard
;   TITLE OF INVENTION: Identification of Essential Genes in
;   TITLE OF INVENTION: Prokaryotes
;   FILE REFERENCE: ELITRA.011A
;   CURRENT APPLICATION NUMBER: US/09/815,242
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 14110
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6700
;   LENGTH: 2820
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2820)
US-09-815-242-6700

Alignment Scores:
Pred. No.:   16   Length: 2820
Score: 56.00   Matches: 11
Percent Similarity: 54.55%   Conservative: 7
Best Local Similarity: 33.33%   Mismatches: 15
Query Match: 24.67%   Indels: 0
DB: 10   Gaps: 0

US-09-833-017B-2 (1-46) x US-09-815-242-6700 (1-2820)

Qy   5 LeuSerLeuLysAsnAspPheLysGluLysThrAspGluLeuGluIleIleIleGly 24
           :::   :::||||:::: |||:::||||:||||:::
Db   988 ATTGATATGGATACGCCGTTTGAAGAATTACCTGCAGACCAACAAGAAATTATTTAAAT 1047

Qy   25 GlySerGlySerLeuSerThrPheArgLeuPheAsn 37
           |||||||||| ::: ||| |||
Db 1048 GGTTCTGGCGGAGAAAAACTTCATTTCATTATGAAAAT 1086

RESULT 9
US-09-070-927A-132
; Sequence 132, Application US/09070927A
; Patent No. US2002012016A1
; GENERAL INFORMATION:
;   APPLICANT: Charles A. Kunsch
;   Patrick J. Dillon
;   Steven Barash
;   TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
;   NUMBER OF SEQUENCES: 982
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/070,927A
;   FILING DATE: 04-May-2000
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/046,655
;   FILING DATE: 1997-05-16
;   APPLICATION NUMBER: 60/044,031
;   FILING DATE: 1997-05-06
;   APPLICATION NUMBER: 60/066,009
;   FILING DATE: 1997-11-14
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Kenley K. Hoover
;   REGISTRATION NUMBER: 40,302
;   REFERENCE/DOCKET NUMBER: PB369
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 132:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5480 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 132:
US-09-070-927A-132

Alignment Scores:
Pred. No.:   39.9   Length: 5480
Score: 56.00   Matches: 11
Percent Similarity: 54.55%   Conservative: 7
Best Local Similarity: 33.33%   Mismatches: 15

```
Query Match:        24.67%      Indels:      0
DB:                 10          Gaps:        0

US-09-833-017B-2 (1-46) x US-09-070-927A-132 (1-5480)

Qy    5 LeuSerLeuLysAsnAspPheLysThrAspGluIleLysThrAspGluLeuGluIleIleGly 24
          :::  ::  ||||:::||||:  |||||:::|:||     ||||||:::
Db 3285 ATTGATATGGATACGGCCGTTTGAAGAATTACCTGCAGACCAACAAGAAAATTATTTTAAAT 3344

Qy   25 GlySerGlySerLeuSerThrPheArgLeuPheAsn 37
          ||||||||       :::  ||||||     |||
Db 3345 GGTTCTGGCGGAGAAAAACTTTCATTTTCATTATGAAAAT 3383

RESULT 10
US-09-933-267A-1/c
; Sequence 1, Application US/09933267A
; Patent No. US2002012309S A1
; GENERAL INFORMATION:
;   APPLICANT: Kalush, Francis et al.
;   TITLE OF INVENTION: Estrogen receptor alpha variants and
;   TITLE OF INVENTION: methods of detection thereof
;   FILE REFERENCE: CL000258CI4
;   CURRENT APPLICATION NUMBER: US/09/933,267A
;   CURRENT FILING DATE: 2001-08-21
;   PRIOR APPLICATION NUMBER: 60/160626
;   PRIOR FILING DATE: 1999-10-20
;   PRIOR APPLICATION NUMBER: 60/183756
;   PRIOR FILING DATE: 2000-02-22
;   PRIOR APPLICATION NUMBER: 09/692414
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 09/768184
;   PRIOR FILING DATE: 2001-01-24
;   PRIOR APPLICATION NUMBER: 09/804076
;   PRIOR FILING DATE: 2001-03-13
;   PRIOR APPLICATION NUMBER: 09/826314
;   PRIOR FILING DATE: 2001-04-05
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 465237
;   TYPE: DNA
;   ORGANISM: human
US-09-933-267A-1

Alignment Scores:
Pred. No.:             1.81e+04      Length:        465237
Score:                 56.00         Matches:       12
Percent Similarity:    57.50%        Conservative:  11
Best Local Similarity: 30.00%        Mismatches:    17
Query Match:           24.67%        Indels:        0
DB:                    10            Gaps:          0

US-09-833-017B-2 (1-46) x US-09-933-267A-1 (1-465237)

Qy    3 LysThrLeuSerLeuLysAsnAspPheLysGluLysLysThrAspLeuGluGluIleIle 22
          ::||||||::|||       |||          :::|||        :::|||
Db 436741 AGAACATTGACAACACCAAATGCTGATGAGGATGGAGCAACAGGAACTCTCACTCATT 436682

Qy   23 IleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGln 42
          ::||||||:|||          :::  ||
Db 436681 GTTGGTGGAAATGGAAAATGGTACAGCCACTGGGAAGATGGTTTGGCAGTTTCCTTCAA 436622

RESULT 11
US-09-881-457A-1
; Sequence 1, Application US/09881457A
; Patent No. US20020081316A1
; GENERAL INFORMATION:
;   APPLICANT: Cochran, Mark D
;   APPLICANT: Cook, Stephanie M
;   APPLICANT: Wild, Martha A
;   TITLE OF INVENTION: No. US2002008131GAlel Avian Herpes Virus and Uses Thereof
;   FILE REFERENCE: SY01105KlQKQK
;   CURRENT APPLICATION NUMBER: US/09/881,457A
```

```
;   CURRENT FILING DATE: 2001-06-14
;   PRIOR APPLICATION NUMBER: 09/426,352
;   PRIOR FILING DATE: 1999-10-25
;   PRIOR APPLICATION NUMBER: 08/804,372
;   PRIOR FILING DATE: 1997-02-21
;   PRIOR APPLICATION NUMBER: PCT/US95/10245
;   PRIOR FILING DATE: 1995-08-09
;   PRIOR APPLICATION NUMBER: 08/663,566
;   PRIOR FILING DATE: 1996-06-13
;   PRIOR APPLICATION NUMBER: 08/288,065
;   PRIOR FILING DATE: 1994-08-09
;   PRIOR APPLICATION NUMBER: PCT/US93/05681
;   PRIOR FILING DATE: 1993-06-14
;   PRIOR APPLICATION NUMBER: 08/023,610
;   PRIOR FILING DATE: 1993-02-26
;   PRIOR APPLICATION NUMBER: 07/898,087
;   PRIOR FILING DATE: 1992-06-12
;   NUMBER OF SEQ ID NOS: 5
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 3570
;   TYPE: DNA
;   ORGANISM: Newcastle disease virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1194)..(2888)
;   OTHER INFORMATION: NDV Fusion Protein
;   NAME/KEY: misc_feature
;   LOCATION: (1355)
;   OTHER INFORMATION: n = any nucleotide
US-09-881-457A-1

Alignment Scores:
Pred. No.:                          Length:        3570
Score:                 26.7         Matches:       16
Percent Similarity:    55.50        Conservative:  7
Best Local Similarity:  47.92%      Mismatches:    20
Query Match:           33.33%       Indels:        5
DB:                    24.45%       Gaps:          1
                       10

US-09-833-017B-2 (1-46) x US-09-881-457A-1 (1-3570)

Qy    4 ThrLeuSerLeuLysAsnAspPhe------------LysGluIleLysThrAspGlu 18
          ||||| |||      ::|||                          |||
Db 2523 ACTTTAAGGCTCAGTGGGAATTCGGATGTAACTTATCAGAAGAATATTCAATACAAGAT 2582

Qy   19 LeuGluIleIleGlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArg 38
          :::|:||||| |||::::   ::::|||||||   ::|||||
Db 2583 TCTCAAGTAATAATAACAGGCAATCTTGATATCTCAACTGAGCTTGGGAATGTCAACAAC 2642

Qy   39 SerPheThrGlnAlaLeuGlyGlyLys 46
          |||     ::    ||||||| |||
Db 2643 TCGATCAGTAATGCCTTGAATAAG 2666

RESULT 12
US-09-788-657-5/c
; Sequence 5, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
;   APPLICANT: Nicolaides, Nicholas
;   APPLICANT: Sass, Philip
;   APPLICANT: Kinzler, Kenneth
;   APPLICANT: Grasso, Luigi
;   APPLICANT: Vogelstein, Bert
;   TITLE OF INVENTION: Methods for generating hypermutable
;   TITLE OF INVENTION: yeast
;   FILE REFERENCE: 01107.00097
;   CURRENT APPLICATION NUMBER: US/09/788,657
;   CURRENT FILING DATE: 2001-02-21
;   PRIOR APPLICATION NUMBER: 60/184,336
;   PRIOR FILING DATE: 2000-02-23
;   NUMBER OF SEQ ID NOS: 25
;   SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 5
; LENGTH: 3218
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-788-657-5

Alignment Scores:
Pred. No.:                     28           Length:              3218
Score:                         55.00        Matches:             15
Percent Similarity:            55.26%       Conservative:        6
Best Local Similarity:         39.47%       Mismatches:          15
Query Match:                   24.23%       Indels:              2
DB:                                         Gaps:                1

US-09-833-017b-2 (1-46) x US-09-788-657-5 (1-3218)

Qy    7 LeuLysAsnAspPheLys------GluIleLysThrAspGluLeuGluIleIleIleGly 24
          ||::||:|||||        ||     ||||:|| ||||::|::    :::|||
Db 2889 CTACAAAATGATATTAGAGATAATTGACTTGAAAAATGATGAAATGTTTTCTCATAAAA 2830

Qy   25 GlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 42
          |||||||||||          ||               :::    ||
Db 2829 GGACTCGGGTCTTTGGTACCGTTGAATAAGCTTAGACAAGAAATCGTTGCTTGAG 2776


RESULT 13
US-09-070-927A-171
; Sequence 171, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
;  APPLICANT: Charles A. Kunsch
;              Patrick J. Dillon
;              Steven Barash
;  TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
;  NUMBER OF SEQUENCES: 982
;  CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Human Genome Sciences, Inc.
;     STREET: 9410 Key West Avenue
;     CITY: Rockville
;     STATE: Maryland
;     COUNTRY: USA
;     ZIP: 20850
;  COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;     COMPUTER: HP Vectra 486/33
;     OPERATING SYSTEM: MSDOS version 6.2
;     SOFTWARE: ASCII Text
;  CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/070,927A
;     FILING DATE: 04-May-2000
;     CLASSIFICATION: <Unknown>
;  PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/046,655
;     FILING DATE: 1997-05-16
;     APPLICATION NUMBER: 60/044,031
;     FILING DATE: 1997-05-06
;     APPLICATION NUMBER: 60/066,009
;     FILING DATE: 1997-11-14
;  ATTORNEY/AGENT INFORMATION:
;     NAME: Kenley K. Hoover
;     REGISTRATION NUMBER: 40,302
;     REFERENCE/DOCKET NUMBER: PB369
;  TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (301) 309-8504
;     TELEFAX: (301) 309-8512
;  INFORMATION FOR SEQ ID NO: 171:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 11871 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: double
;        TOPOLOGY: linear
;     SEQUENCE DESCRIPTION: SEQ ID NO: 171:
US-09-070-927A-171
```

```
Alignment Scores:
Pred. No.:                     206          Length:              11871
Score:                         54.50        Matches:             17
Percent Similarity:            53.49%       Conservative:        6
Best Local Similarity:         39.53%       Mismatches:          15
Query Match:                   24.01%       Indels:              5
DB:                                         Gaps:                2

US-09-833-017b-2 (1-46) x US-09-070-927A-171 (1-11871)

Qy    1 MetLysLys---------ThrLeuSerLeuLysAsnAspPheLysGluIleLysThr 16
          |||||||||        |||      :::::||:||    |||||
Db 4024 ATGAAGAAATATTCGAATGGTACAACAGCGATTCGGAATATTCAGTAGAGATTGACCAA 4083

Qy   17 AspGluLeuGluIleIleIleGlyGlySerGlySer---LeuSerThrPheArgLeu 35
           ||       |||           :|||||:::|        ||||||||
Db 4084 GGCGAGTTCGTCTATGTGGTTGGTCCCTCAGGTGCAGGGAAATCAACCTTTATTAGATTA 4143

Qy   36 PheAsnArg 38
           ||
Db 4144 ATGTATCGT 4152


RESULT 14
US-09-917-800A-1496
; Sequence 1496, Application US2002011946A1
; Patent NO. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT: Porter, Mark
;  APPLICANT: Johnson, Kory
;  APPLICANT: Castle, Arthur
;  APPLICANT: Elashoff, Michael
;  APPLICANT: Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
;  SEQ ID NO 1496
;     LENGTH: 2376
;     TYPE: DNA
;     ORGANISM: Rattus norvegicus
;     FEATURE:
;     OTHER INFORMATION: Genbank Accession No. US20020119462A1 X51615
US-09-917-800A-1496

Alignment Scores:
Pred. No.:                     27           Length:              2376
Score:                         54.00        Matches:             21
Percent Similarity:            43.55%       Conservative:        6
Best Local Similarity:         33.87%       Mismatches:          9
Query Match:                   23.79%       Indels:              26
DB:                                         Gaps:                4
```

US-09-833-017B-2 (1-46) x US-09-917-800A-1496 (1-2376)

```
Qy     7 LeuLysAsnAspPheLys--------GluIleLysThrAspGluLeuGluIleIleIle 23
         :::||||||:::||||||       |||||||||||   :::::||      |||
Db   605 ATAAAGAACCAGTTTAAGGACATCGAAGAGATCAAAACCCAGAAGGTCCGTATC------ 658

Qy    24 GlyGlySerGlySerLeu---------------SerThrPhePheArg---------- 34
         |||||||||                       |||  ||||||||||
Db   659 ------GAAGGGTCCCTGTGGTGGACCTACACCACCAGCATCTTCTTCCGGGTCATCTTC 712

Qy    35 --LeuPheAsnArgSerPheThrGlnAlaLeu 44
           :::::|||        ||| ||| |||
Db   713 GAAGCTGTCTTCATCTATGTCTTTTACATCATGTTACATGGCTTCTTCATGCAGCGTCTG 772

Qy    45 GlyLys 46
         |||
Db   773 GTGAAG 778
```

RESULT 15
US-09-973-451-38
; Sequence 38, Application US/09973451
; Patent No. US2002013228A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
;   LENGTH: 29793
;   TYPE: DNA
;   ORGANISM: Caenorhabaditis elegans
;   FEATURE:
US-09-973-451-38

Alignment Scores:
Pred. No.:   887
Score: 54.00
Percent Similarity: 61.11%
Best Local Similarity: 47.22%
Query Match: 23.79%
DB: 10

Length: 29793
Matches: 17
Conservative: 5
Mismatches: 12
Indels: 2
Gaps: 2

US-09-833-017B-2 (1-46) x US-09-973-451-38 (1-29793)

```
Qy     9 AsnAspPheLysGluIleLysThrAspGlu---LeuGluIleIleIleGlyGlySerGly 27
          ||||||     ::  :::::::  |||         ||||||||||||       |||
Db  2356 AATGATCTACAAATCGTTCAATTGGATCCTGTCTTGGAAATAATAATTTCCAATTCGTAA 2415

Qy    28 SerLeu---SerThrPhePheArgLeuPheAsnArgSerPheThrGln 42
          ||||||   ||||||| |||  ::: |||||| ::: |||
Db  2416 AGTTTGCATTCAACTCAATTTCTCGTTTCAAATCGTCGATATCCGCAA 2463
```

Search completed: November 9, 2002, 00:58:43
Job time : 98.3134 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: November 8, 2002, 22:01:57 ; Search time 1808.42 Seconds
(without alignments)
411.958 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 227
Sequence: 1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext 7.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105349_4846/app-query.fasta_1.398
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=0 -2000000000
-USER=US09833017_@CGN_1_1_899_@runat_05112002_105349_4846 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
          1:   em_estba:*
          2:   em_esthum:*
          3:   em_estin:*
          4:   em_estmu:*
          5:   em_estov:*
          6:   em_estpl:*
          7:   em_estro:*
          8:   em_htc:*
          9:   gb_est1:*
         10:   gb_est2:*
         11:   gb_htc:*
         12:   gb_est3:*
         13:   gb_est4:*
         14:   gb_est5:*
         15:   em_estfun:*
         16:   em_estom:*
         17:   gb_gss:*
         18:   em_gss_hum:*
         19:   em_gss_inv:*
         20:   em_gss_pln:*
         21:   em_gss_vrt:*
         22:   em_gss_fun:*
         23:   em_gss_mam:*
         24:   em_gss_mus:*
         25:   em_gss_other:*
         26:   em_gss_pro:*
         27:   em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		67	29.5	691	13	BJ069595	BJ069595 BJ069595
2		63	27.8	328	9	AA358436	AA358436 EST67334
3		62	27.3	527	13	BI707485	BI707485 fs41f04.y
4		62	27.3	565	13	AW547369	AW547369 L001C10-
5		62	27.3	621	13	BI841174	BI841174 fq52b06.y
6		62	27.3	720	13	BI891813	BI891813 ZF637-3-0
7		62	27.3	721	12	BG587685	BG587685 EST489460
8	c	62	27.3	1142	17	CNS06RAR	AL411721 T7 end of
9	c	62	27.3	1391	12	BF207257	BF207257 601870728
10	c	61.5	27.1	661	12	AV321480	AV321480 AV321480
11	c	61.5	27.1	818	12	BG678629	BG678629 602624419
12	c	61	26.9	250	10	BB067911	BB067911 BB067911
13	c	61	26.9	906	17	A2528065	A2528065 ENTBT05TR
14		60.5	26.7	560	17	A2849249	A2849249 2M0150B07
15		60	26.4	483	17	A2149972	A2149972 SP_0037_B
16	c	60	26.4	489	12	BF435651	BF435651 nac34f12.
17	c	60	26.4	636	17	BH603995	BH603995 BOGQ193TF
18		60	26.4	808	17	CNS04WJF	AL310452 Tetraodon
19		60	26.4	1118	17	BF384061	BF384061 60045592
20	c	59.5	26.2	634	10	BB375066	BB375066 BB375066
21	c	59	26.0	302	10	AV340411	AV340411 AV340411
22		59	26.0	394	9	AI553452	AI553452 vw55f10.x
23		59	26.0	506	10	AW861203	AW861203 RC1-CT030
24	c	59	26.0	609	10	AW254713	AW254713 ML100_pep
25	c	59	26.0	686	14	B0007240	B0007240 QGH13N12.
26	c	59	25.8	436	10	BB819813	BB819813 BB819813
27		58.5	25.8	624	17	BH385411	BH385411 AG-ND-146
28		58.5	25.8	661	12	BE201479	BE201479 fk93d08.x
29	c	58.5	25.8	789	12	BE877229	BE877229 60148S150
30	c	58	25.6	252	17	BH770061	BH770061 BMBAC369F
31	c	58	25.6	432	17	AQ085216	AQ085216 60231262Z
32	c	58	25.6	609	17	BH404546	BH404546 AG-ND-137
33	c	58	25.6	624	9	AL507817	AL507817 AL507817
34	c	58	25.6	658	17	BB369772	BB369772 BB369772
35		58	25.6	744	12	BG597482	BG597482 EST496160
36		57.5	25.3	363	12	BG055682	BG055682 naf62h04.
37		57.5	25.3	539	10	AV526794	AV526794 AV526794
38		57.5	25.3	562	13	BG952882	BG952882 PM4-CT054
39	c	57.5	25.3	615	12	BG661558	BG661558 kx03d05.y
40		57.5	25.3	676	12	AV822067	AV822067 AV822067
41	c	57.5	25.3	762	17	AQ157280	AQ157280 nbxb00090
42		57.5	25.3	869	12	BG105781	BG105781 602312622
43	c	57.5	25.3	954	17	CNS02J1B	AL199640 Tetraodon
44	c	57	25.1	294	9	AI986937	AI986937 rs19d12.y
45	c	57	25.1	333	10	BB095765	BB095765 BB095765

ALIGNMENTS

RESULT 1
BJ069595
LOCUS   BJ069595   691 bp   mRNA   linear   EST 11-DEC-2001
DEFINITION BJ069595 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL053o10 5', mRNA sequence.
ACCESSION   BJ069595
VERSION   BJ069595.1 GI:17497955
KEYWORDS   EST.
SOURCE   African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 691)
AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara

```
TITLE        ,Y.
             Expressed genes in X. laevis embryo
JOURNAL      Unpublished (2001)
COMMENT      Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES     Location/Qualifiers
     source  1..691
             /organism="Xenopus laevis"
             /db_xref="taxon:8355"
             /clone_lib="NIBB Mochii normalized Xenopus tailbud
             library"
             /tissue_type="whole embryo"
             /dev_stage="stage 25"
             /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
             were oligo-dT primed and directionally cloned. Staging
             according to Nieuwkoop and Faber. Library is subtracted
             and was constructed by N. Garrett and A.M. Zorn,
             (Wellcome/CRC Institute)."

BASE COUNT    173 a    148 c    206 g    162 t    2 others
ORIGIN

Alignment Scores:
Pred. No.:               3.21          Length:            691
Score:                  67.00          Matches:            12
Percent Similarity:     64.52%         Conservative:        8
Best Local Similarity:  38.71%         Mismatches:         11
Query Match:            29.52%         Indels:              0
DB:                                    Gaps:                0

US-09-833-017B-2 (1-46) x BJ069595 (1-691)

Qy    12 LysGluIleLysThrAspPheAspGluGluGluIleIleIleGlyGlySerGlySerLeuSerThr 31
         ::|::||||||||   ::|  |||||||||   |||:||
Db   451 CGGCAGGTTGGAGTGGACATAAATGAAAACTGATTGCCAGGCAGCAGCTCGATTGAAGAA 510

Qy    32 PhePheArgLeuPheAsnArgSerPheThrGln 42
         |||||||||| ||||||:||| |||||||:|::
Db   511 TTCTTCCGACTCATGAATCGGCAATTCACGGAG 543
```

```
RESULT 2
AA358436
LOCUS       AA358436
DEFINITION  EST67334 Fetal lung III Homo sapiens cDNA 5' end similar to EST
            containing L1 repeat, mRNA sequence.
ACCESSION   AA358436
VERSION     AA358436.1  GI:2010774
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
AUTHORS     Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE       Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
JOURNAL     Nat. Genet. 4, 373-380 (1993)
MEDLINE     94004965
            Other_ESTs: EST67333
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..328
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):162615"
                     /db_xref="taxon:9606"
                     /clone_lib="Fetal lung III"
                     /dev_stage="fetus"
                     /note="Organ:lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI"

BASE COUNT    145 a    60 c    58 g    59 t    6 others
ORIGIN

Alignment Scores:
Pred. No.:               5.32          Length:            328
Score:                  63.00          Matches:            17
Percent Similarity:     51.22%         Conservative:        4
Best Local Similarity:  41.46%         Mismatches:         14
Query Match:            27.75%         Indels:              6
DB:                                    Gaps:                1

US-09-833-017B-2 (1-46) x AA358436 (1-328)

Qy     6 SerLeuLysAsnAspPheLysGluGluGluIleIleIleGlyGly 25
         :::::|||||||||    ::|||  ||||||  |||||
Db   178 GCAATAAAAATGATAAAGGGATATCACCACCGATCCCACAGAAATACAAACTACCNTC 237

Qy    26 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaAlaLeuGly 45
         |||||||      ---------------  |||||||  ||||||
Db   238 AGAGGGATA---------------CTTTTCAACANCTCTTACACAAATNAACTNGGG 279

Qy    46 Lys 46
         |||
Db   280 AAA 282
```

```
RESULT 3
BI707485
LOCUS       BI707485                 527 bp    mRNA    linear    EST 19-SEP-2001
DEFINITION  fs41f04.y1 zebrafish adult olfactory Danio rerio cDNA clone 507046
            5' similar to SW:PTB_HUMAN P26599 POLYPYRIMIDINE TRACT-BINDING
            PROTEIN ;, mRNA sequence.
ACCESSION   BI707485
VERSION     BI707485.1  GI:15683180
KEYWORDS    EST.
SOURCE      zebrafish.
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 527)
AUTHORS     Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE       WashU Zebrafish EST Project 1998
JOURNAL     Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: John Ngai cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
```

```
               Seq primer: T3 ET from Amersham
               High quality sequence stop: 428.
FEATURES       Location/Qualifiers
     source    1..527
               /organism="Danio rerio"
               /db_xref="taxon:7955"
               /clone="5070463"
               /clone_lib="Zebrafish adult olfactory"
               /sex="mixed"
               /tissue_type="Olfactory rosettes"
               /dev_stage="adult"
               /lab_host="DI0Hb (Gibco BRL)"
               /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
               is a directionally cloned cDNA library from adult
               zebrafish olfactory epithelium."
BASE COUNT     147 a    137 c    101 g    142 t
ORIGIN
```

Alignment Scores:
Pred. No.:   14.2   Length: 527
Score: 62.00   Matches: 14
Percent Similarity: 53.66%   Conservative: 8
Best Local Similarity: 34.15%   Mismatches: 19
Query Match: 27.31%   Indels: 0
DB: 13   Gaps: 0

US-09-833-017B-2 (1-46) x BI707485 (1-527)

```
Qy    4 ThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIleIle 23
         ||||||  |||  |||            ::       :::::::::::::
Db  196 ACTCTGCACCTGCTAATATCCCACCCTCGTTCGTTGAGGATGACCTTAAGCTGCTCTT 255

Qy   24 GlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAla 43
         |||  ||||||:::|||          |||          |||
Db  256 GGTAGCAGTGGTGCTCTGGTCAAGAACTTTAAGTTCTTTCAAAAAGATCGCAAGATGGCT 315

Qy   44 Leu 44
         |||
Db  316 CTG 318
```

RESULT 4
AW547369
```
LOCUS       AW547369         565 bp    mRNA       linear   EST 31-AUG-2000
DEFINITION  L0021C10-3 NIA Mouse E12.5 Female Mesonephros and Gonads cDNA
            Library Mus musculus cDNA clone L0021C10 3', mRNA sequence.
ACCESSION   AW547369
VERSION     AW547369.1  GI:7192783
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
            ,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
            III, Becker,K.G. and Ko,M.S.H.
  TITLE     Genome-wide expression profiling of mid-gestation placenta and
            embryo using a 15,000 mouse developmental cDNA microarray
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
  MEDLINE   20381348
  COMMENT   Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: L0021 row: C column: 10
            Seq primer: -21M13 Forward
            High quality sequence stop: 565
            POLYA=Yes.
FEATURES       Location/Qualifiers
     source    1..565
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="niaEST:L0021C10-3"
               /db_xref="taxon:10090"
               /clone="L0021C10"
               /clone_lib="NIA Mouse E12.5 Female Mesonephros and Gonads
               cDNA Library"
               /sex="female"
               /dev_stage="12.5dpc"
               /lab_host="DH10B"
               /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
               Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
               2 Mesonephros. The double-stranded cDNA was synthesized
               by Gibco's kit with an Oligo(dT) primer [NotI
               primer-adapter from GibcoBRL]
               [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
               3.42ug of total RNA. The double-stranded cDNAs were
               treated with T4 DNA polymerase and purified by
               ethanol-precipitation. The cDNAs were ligated to
               Lone-linker LL-Sal3 (include SalI sequence). The cDNAs
               were purified by phenol/chloroform and separated from
               free linkers by Centricon 100. Then, cDNAs were amplified
               by long-range high fidelity PCR using Takara's Ex Taq
               polymerase. Then, the cDNAs were purified by
               phenol/chloroform and by Centricon 100. The cDNAs were
               digested with SalI and NotI enzymes. Then, the cDNAs were
               size selected by Gibco's Size Fractionation Column. The
               cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
               vector. The DH10B E. coli host was transformed with the
               ligation mixture by chemical method. The library was
               constructed by Xiaohong Wang."
BASE COUNT     92 a    153 c    143 g    177 t
ORIGIN
```

Alignment Scores:
Pred. No.:   15.5   Length: 565
Score: 62.00   Matches: 18
Percent Similarity: 59.52%   Conservative: 7
Best Local Similarity: 42.86%   Mismatches: 11
Query Match: 27.31%   Indels: 6
DB: 10   Gaps: 2

US-09-833-017B-2 (1-46) x AW547369 (1-565)

```
Qy    5 LeuSerLeuLysAsnAspPheLysGluIleLysThrAsp------------GluLeu 19
         ||| |||  |||  :::       ||  :::::::::|||||
Db  236 TTGGATTTACTGTCCCCAGCAAAGGCCCTGAAGTCTGACTGGCTGAGTCATAGGAACTG 295

Qy   20 GlyIleIleIleGlyLysGlySerLeuSerThrPheArgLeuPheAsnArgSer 39
         ||||||||||||||||||||::::|||     |||  |||||||:::|||
Db  296 AAGTCTTTCCCTGGGGATCAGGTCCTTGGCT---TTCTTGGAGCTTTTAGTCGTTTC 352

Qy   40 PheThr 41
         ||||:|
Db  353 TTCTCC 358
```

RESULT 5
BI841174
```
LOCUS       BI841174        621 bp    mRNA       linear   EST 04-OCT-2001
DEFINITION  fq52b06.y1 Zebrafish adult olfactory Danio rerio cDNA clone 4879450
            5' similar to SW:PTB_HUMAN P26599 POLYPYRIMIDINE TRACT-BINDING
            PROTEIN ;, mRNA sequence.
ACCESSION   BI841174
VERSION     BI841174.1  GI:15953697
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
```

TITLE   Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
JOURNAL   WashU Zebrafish EST Project 1998
            Unpublished (1998)
COMMENT   Other_ESTs: fq52b06.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: John Ngai cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 479.
FEATURES   Location/Qualifiers
     source   1..621
                    /organism="Danio rerio"
                    /db_xref="taxon:7955"
                    /clone="4879450"
                    /clone_lib="zebrafish adult olfactory"
                    /sex="mixed"
                    /tissue_type="Olfactory rosettes"
                    /dev_stage="adult"
                    /lab_host="D10Hb (Gibco BRL)"
                    /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                    is a directionally cloned cDNA library from adult
                    Zebrafish olfactory epithelium."
BASE COUNT   170 a   150 c   115 g   186 t
ORIGIN

Alignment Scores:
Pred. No.:   17.5   Length: 621
Score: 62.00   Matches: 14
Percent Similarity: 53.66%   Conservative: 8
Best Local Similarity: 34.15%   Mismatches: 19
Query Match: 27.31%   Indels: 0
DB: 13   Gaps: 0

US-09-833-017B-2 (1-46) x BI841174 (1-621)

Qy   4 ThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIleLeuIle 23
           ||||| || ||| ::: |||:::|||::|||::|||::||:
Db   198 ACTCTGCACCTGTCTAATATCCCACCCTCGTTGTTGAGGATGACCTTAAGCTGCTCTTT 257

Qy   24 GlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAla 43
           ||| |||||||:::||| :|| ||| ||: :::
Db   258 GGTAGCAGTGGTGCTCGTCAAGAACTTTAAGTCTTTCAAAAAGATCGCAAGATGGCT 317

Qy   44 Leu 44
           |||
Db   318 CTG 320

RESULT 6
BI891813
LOCUS   BI891813   720 bp   mRNA   linear   EST 12-OCT-2001
DEFINITION 2F637-3-001068 zebrafish shield stage whole embryo cDNA library
            MPMGp637 Danio rerio cDNA clone MPMGp637_19A6:MPMGp637A0619 5',
            mRNA sequence.
ACCESSION   BI891813
VERSION   BI891813.1 GI:16099084
KEYWORDS   EST.
SOURCE   Danio rerio
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.

REFERENCE   1 (bases 1 to 720)
AUTHORS   Clark,M., Aanstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.
TITLE   EST sequencing of a zebrafish shield stage cDNA library normalised
            by oligonucleotide fingerprinting
JOURNAL   Unpublished (2001)
COMMENT   Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            5' EST sequencing of clones from a zebrafish shield stage library,
            normalised from 55,000 starting clones by oligonucleotide
            fingerprinting
            High quality sequence stop: 720.
FEATURES   Location/Qualifiers
     source   1..720
                    /organism="Danio rerio"
                    /db_xref="taxon:7955"
                    /clone="MPMGp637_19A6:MPMGp637A0619"
                    /clone_lib="zebrafish shield stage whole embryo cDNA
                    library MPMGp637"
                    /tissue_type="whole embryo"
                    /dev_stage="shield stage, 6 hrs post-fertilisation"
                    /lab_host="E.coli, XL1 blue MRF"
                    /note="Vector: pSportl; Site_1: NotI; Site_2: SalI;
                    oligo-dT-NotI primed, SalI adaptors, directionally cloned,
                    library normalised by oligonucleotide fingerprinting"
BASE COUNT   199 a   169 c   138 g   208 t   6 others
ORIGIN

Alignment Scores:
Pred. No.:   21.2   Length: 720
Score: 62.00   Matches: 14
Percent Similarity: 53.66%   Conservative: 8
Best Local Similarity: 34.15%   Mismatches: 19
Query Match: 27.31%   Indels: 0
DB: 13   Gaps: 0

US-09-833-017B-2 (1-46) x BI891813 (1-720)

Qy   4 ThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIleLeuIle 23
           ||||| || ||| ::: |||:::|||::|||::||:
Db   292 ACTCTGCACCTGTCTAATATCCCACCCTCGTTGTTGAGGATGACCTTAAGCTGCTCTTT 351

Qy   24 GlyGlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAla 43
           ||| |||||||:::||| ::: ||| ||| ::: :::
Db   352 GGTAGCAGTGGTGCTCGTCAAGAACTTTAAGTCTTTCAAAAAGATCGCAAGATGGCT 411

Qy   44 Leu 44
           |||
Db   412 CTG 414

RESULT 7
BG587685/c
LOCUS   BG587685   721 bp   mRNA   linear   EST 11-APR-2001
DEFINITION EST489460 MHAM Medicago truncatula/Glomus versiforme mixed EST
            library cDNA clone pMHAM-50F2 5' end, mRNA sequence.
ACCESSION   BG587685
VERSION   BG587685.1 GI:13602749
KEYWORDS   EST.
SOURCE   Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM   Medicago truncatula/Glomus versiforme mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1 (bases 1 to 721)
AUTHORS   Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
            and Fraser,C.M.
TITLE   ESTs from roots of Medicago truncatula after colonization with
            Glomus versiforme, 2001
JOURNAL   Unpublished (2001)
COMMENT   Contact: Harrison M.J.
            Plant Biology Division

The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N38675B TIGR sequence name: MTDDJ25TK More
information is available at: http://www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

```
FEATURES             Location/Qualifiers
     source          1..721
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:l9092"
                     /clone="pMHAM-50F2"
                     /clone_lib="MHAM"
                     /tissue_type="roots colonized with Glomus versiforme"
                     /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
                     directionally ligated into the Unizap XR vector from
                     stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT       219 a    143 c    164 g    195 t
ORIGIN
```

Alignment Scores:
Pred. No.:   21.3   Length: 721
Score: 62.00   Matches: 17
Percent Similarity: 50.98%   Conservative: 9
Best Local Similarity: 33.33%   Mismatches: 9
Query Match: 27.31%   Indels: 16
DB: 12   Gaps: 2

```
US-09-833-017B-2 (1-46) x BG587685 (1-721)

Qy   6  SerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGlu----------- 20
           |||  ::::::||::::::||:::::||:||:::||
Db 671  TCTGCGATGACAGATCCCAAGGAAGTGGAATCAGGAGGAACTGGAATTGATCAACTCCG 612

Qy  21  ----------IleIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeu 35
                       :::::   ||   :   :::::||||
Db 611  GTGGTGGTGGTGGTGGTTGTGGCTGGTGCTTCCCGCGCTTCCGAGTTTTTTTTTTTT 552

Qy  36  ------------PheAsnArgSerPhe 40
                    ||||||||||||
Db 551  TTTTTTAGTGCTTAAACTTTTAACCGTAGTTTT 519
```

RESULT 8
CNS06RAR/c
LOCUS   CNS06RAR   1142 bp   DNA   linear   GSS 05-JUL-2001
DEFINITION T7 end of clone AW0AA012H05 of library AW0AA from strain CLIB 89 of
            Yarrowia lipolytica, genomic survey sequence.
ACCESSION   AL411721
VERSION   AL411721.1 GI:12181438
KEYWORDS   GSS.
SOURCE   Yarrowia lipolytica
ORGANISM   Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1 (bases 1 to 1142)
AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

TITLE   Wincker,P. and Weissenbach,J.
            Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE   20584711
PUBMED   11152876
REFERENCE   2 (bases 1 to 1142)
AUTHORS   Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE   Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
            lipolytica
JOURNAL   FEBS Lett. 487 (1), 95-100 (2000)
MEDLINE   20584727
PUBMED   11152892
REFERENCE   3 (bases 1 to 1142)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.

```
FEATURES             Location/Qualifiers
     source          1..1142
                     /organism="Yarrowia lipolytica"
                     /strain="CLIB 89"
                     /db_xref="taxon:4952"
                     /clone="AW0AA012H05"
                     /clone_lib="AW0AA"
                     /note="end : T7"
     misc_feature    <108..>1116
                     /note="similar to Saccharomyces cerevisiae ORF YHR047c [
                     AAP1 ; alanine/arginine aminopeptidase ]
                     1 putative frameshift(s)
                     similar to Saccharomyces cerevisiae ORF YKL157w [ APE2 ;
                     aminopeptidase yscII ]"
                     /evidence=not_experimental
BASE COUNT       278 a    320 c    267 g    273 t
ORIGIN
```

Alignment Scores:
Pred. No.:   38.5   Length: 1142
Score: 62.00   Matches: 15
Percent Similarity: 48.84%   Conservative: 6
Best Local Similarity: 34.88%   Mismatches: 16
Query Match: 27.31%   Indels: 6
DB: 17   Gaps: 1

```
US-09-833-017B-2 (1-46) x CNS06RAR (1-1142)

Qy   9  AsnAspPheLysGluIleLysThrAspGluIleIleIleGlyLeuGluIleIleGlyGlySerGlySer 28
           |||::::||:||:::::|||  ||:::||   :::|||   |||   ||:||
Db 356  AATAATCTTGGGAACTTAAACGTGACCCGTTTCGGCATCTTTGTCGTAGGTGACGGTCGC 297

Qy  29  LeuSerThrPhePheArg----------------------LeuPheAsnArgSerPheThrGln 42
           |||||   ||   :||                       |:::|||   :::||:
Db 296  CTCTCCAATTTCTTCAATGCAACTCGGGAGATCTCCTGATCAATCGAGTTGACAGTGAG 237

Qy  43  AlaLeuGly 45
           :::||||
Db 236  GGTGTCGGT 228
```

RESULT 9
BF207257
LOCUS   BF207257   1391 bp   mRNA   linear   EST 06-NOV-2000

DEFINITION   6018T0728F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4100790 5',
              mRNA sequence.
ACCESSION   BF207257
VERSION   BF207257.1 GI:11100843
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1391)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapps-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCM971 row: h column: 07
              High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..1391
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4100790"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming; Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   409 a   357 c   292 g   326 t   7 others
ORIGIN

Alignment Scores:
Pred. No.:   49.7
Score: 62.00   Length: 1391
Percent Similarity: 53.49%   Matches: 14
Best Local Similarity: 32.56%   Conservative: 9
Query Match: 27.31%   Mismatches: 20
DB: 12   Indels: 0
                                                Gaps: 0

US-09-833-017b-2 (1-46) x BF207257 (1-1391)

Qy   4 ThrLeuSerLeuLysAsnAspPheLysGluIleLysThrAspGluLeuGluIleIleIle 23
        ::|||:||||||||:|::   :|| :::|||
Db 632 TCTTTGGCACTAAGGCTCCCAACCCTTTTCTGAAGTGCTCTCAAGCGCACCGGTTAATC 691

Qy   24 GlyGlySerGlySerLeuSerPheThrPheArgLeuPheAsnArgSerPheThrGlnAla 43
        ||::: |||||||||| ||||| ||:::|
Db 692 AACGGAACTCAGAGGTTATCAACTCTCTTTAGAACTTTCCACCGTGGCCTGGATATGACT 751

Qy   44 LeuGlyLys 46
        ||||||:::
Db 752 CTTGGTAGA 760

RESULT 10
AV321480/c
LOCUS   AV321480   661 bp   mRNA   linear   EST 24-OCT-2001
DEFINITION AV321480 RIKEN full-length enriched, 13 days embryo male testis Mus
            musculus cDNA clone 6030434H13 3', mRNA sequence.
ACCESSION   AV321480
VERSION   AV321480.2 GI:16395068
KEYWORDS   EST.
SOURCE   house mouse.

ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 661)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito.M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu.M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL   Unpublished (2001)
COMMENT   On Nov 9, 1999 this sequence version replaced gi:6291313.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp.
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu.M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S. Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in RIKEN
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES   Location/Qualifiers
     source   1..661
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6030434H13"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     male testis"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse

```
                transcriptase and subsequently enriched for full-length by
                cap-trapper. cDNA went through one round of normalization
                to Rot = 5.0 and subtraction to Rot = 100.0. Second strand
                cDNA was prepared with the primer adapter of sequence (5'
                GAGAGAGAGATTCCGAGTTAATAAATTAAATTAATCCCCCCCCCC 3']. cDNA
                was cloned into the XhoI and BamHI sites. Vector: a
                modified pBluescript KS(+) after bulk excision from Lambda
                FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       184 a   106 c   142 g   229 t
ORIGIN

Alignment Scores:
Pred. No.:              22.8         Length:            661
Score:                  61.50        Matches:           20
Percent Similarity:     58.00%       Conservative:      9
Best Local Similarity:  40.00%       Mismatches:        14
Query Match:            27.09%       Indels:            7
DB:                                  Gaps:              2

US-09-833-017B-2 (1-46) x AV321480 (1-661)

Qy    2 LysLysThrLeu------SerLeuLysAsnAspPheLysLysLeuThrAspGluLeu 19
          |||||| |||      ||||||||||:::       :::   |||:|||||||
Db  153 AAGAAACAGCTTAAATTAGTTTGAAAAATGAAAGAAAGATAGAGCTGACCAGGTT 94

Qy   20 GluIleIleGlyGlySerGlySerLeuSerThr---------------PhePheArg 34
          ||||||:::||||||||||:::|                       |||  |||
Db   93 GAAATTTGACTTAACATTCGGGAGTATCGATACCCCGAGGAGCACCTGGCAATTCTAG 34

Qy   35 LeuPheAsnArgSerPheThrGlnAlaLeu 44
          ||   :::|||||        ||||||
Db   33 GCCTTTTTCTGGCTATTTCAAAAAGCACTA 4

RESULT 11
BG678629/c
LOCUS       BG678629            818 bp   mRNA      linear   EST 01-MAY-2001
DEFINITION  60262441F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749317 5',
            mRNA sequence.
ACCESSION   BG678629
VERSION     BG678629.1  GI:13910026
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 818)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10602 row: f column: 06
            High quality sequence stop: 818.
FEATURES             Location/Qualifiers
     source          1..818
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4749317"
                     /clone_lib="NCI_CGAP_Skn4"
                     /tissue_type="squamous cell carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       171 a   212 c   194 g   241 t


ORIGIN

Alignment Scores:
Pred. No.:              30.1         Length:            818
Score:                  61.50        Matches:           15
Percent Similarity:     60.98%       Conservative:      10
Best Local Similarity:  36.59%       Mismatches:        15
Query Match:            27.09%       Indels:            1
DB:                                  Gaps:              1

US-09-833-017B-2 (1-46) x BG678629 (1-818)

Qy    2 LysLysThrLeuSerLeuLysAsnAspPheLysLysThrAspGluLeuGluIle 21
          :::|||       ::::|||        ::::        |||||
Db  757 AGGAAGCAGCAGCCCTCGGGTCAGACTCCTATTAGCTTGAACTGGTGACT 698

Qy   22 IleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThr 41
          ||| ::::|||  ||||||       ::::|||    |||:|||   |||:|||
Db  697 TTGCTTGGTTCACTGGGTTAATGAGCTCGGTGTTTAGG---TTTGACCGTGACTTTAGC 641

Qy   42 Gln 42
          |||
Db  640 CAG 638

RESULT 12
BB067911/c
LOCUS       BB067911            250 bp   mRNA      linear   EST 25-JUN-2000
DEFINITION  BB067911 RIKEN full-length enriched, 15 days embryo male testis Mus
            musculus cDNA clone 8030468G23 3', mRNA sequence.
ACCESSION   BB067911
VERSION     BB067911.1  GI:8525325
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 250)
AUTHORS     Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,Y., Tagawa,A., Takahashi,F., Tominaga,T., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL     Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J,
            Tomaru,Y., Carninci,P., Shibata,K., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
```

Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

```
FEATURES             Location/Qualifiers
     source          1..250
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="8030468G23"
                     /clone_lib="RIKEN full-length enriched, 15 days embryo
                     male testis"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="15 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGACGTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 185.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT          58 a          62 c          41 g          89 t
ORIGIN

Alignment Scores:
Pred. No.:                    7.8          Length:         250
Score:                       61.00         Matches:         12
Percent Similarity:          66.67%        Conservative:     6
Best Local Similarity:       44.44%        Mismatches:       9
Query Match:                 26.87%        Indels:           0
DB:                          10            Gaps:             0

US-09-833-017B-2 (1-46) x BB067911 (1-250)

Qy     2 LysLysThrLeuLeuSerLeuLysAsnAspPheLysGluIleLysThrGluLeuGluIle 21
           ::|||||||::. ::||||||   ::||||   |||||  |||   :::||
Db   117 AGGAAACAATGAAATAAAAAATGTACAAAAGTATAAAGAGGGAGAAATTACAATA 58

Qy    22 IleIleGlyGlySerGlySer 28
           |||||||   |||..:
Db    57 TGGAGGGTGGGTGCGGGAAT 37
```

RESULT 13
```
AZ528065/c

LOCUS       AZ528065              906 bp    DNA     linear   GSS 03-NOV-2000
DEFINITION  ENTBT05TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ528065
VERSION     AZ528065.1  GI:11080269
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
```

Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 99
High quality sequence stop: 763.

```
FEATURES             Location/Qualifiers
     source          1..906
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                     method for isolate identification. Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT         311 a         146 c          96 g         353 t
ORIGIN

Alignment Scores:
Pred. No.:                   41.2          Length:          906
Score:                       61.00         Matches:          14
Percent Similarity:          58.97%        Conservative:      9
Best Local Similarity:       35.90%        Mismatches:       14
Query Match:                 26.87%        Indels:            2
DB:                          17            Gaps:              1

US-09-833-017B-2 (1-46) x AZ528065 (1-906)

Qy     8 LysAsnAspPheLysGluIleLysThrAspGluLeuGluIleIleGlyGlySerGly 27
           :::||||||   ::||||  |||   |||||||   :::||||||
Db   730 CAAAACGAGGCCAGAGAATGTAAAGTAAAAAAAATTGAAACATTTGTTGGAGGTAAATG 671

Qy    28 SerLeuSerThrPhe----------PheArgLeuPheAsnArgSerPheThrGlnAlaLeu 44
           ||||||||          |||||||||    :::       |||||::  :::
Db   670 GACCATTCCACATTGCCTGTTTTAGACTTGATGATACCGCTTTTTACTGAAAAATA 614
```

RESULT 14
```
AZ849249

LOCUS       2M0150B07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0150B07 R, DNA sequence.
                           560 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0150B07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0150B07 R, DNA sequence.
ACCESSION   AZ849249
VERSION     AZ849249.1  GI:13033130
KEYWORDS    GSS.
SOURCE      Mus musculus
  ORGANISM  house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
```

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000   Std Error: 0.00
Plate: 0150   row: B   column: 07
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 560.

```
FEATURES      Location/Qualifiers
     source   1..560
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC2M0150B07"
              /clone_lib="Mouse 10kb plasmid UUGC1M library"
              /sex="Male"
              /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptored DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adaptored mouse DNA was annealed to
              adaptored vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
```

BASE COUNT   238 a   102 c   70 g   150 t
ORIGIN

Alignment Scores:
Pred. No.:   26.6   Length: 560
Score: 60.50   Matches: 15
Percent Similarity: 55.32%   Conservative: 11
Best Local Similarity: 31.91%   Mismatches: 18
Query Match: 26.65%   Indels: 3
DB: 17   Gaps: 1

US-09-833-017B-2 (1-46) x AZ849249 (1-560)

```
Qy   1 MetLysLysThrLeuSerLeu-------LysAsnAspPheLysGlyLeuSerLeuLysThrAsp 17
         |||  ::||||  ||::||       |||||  |||    ::||
Db 112 CTGATGCAAACATTCTCAATAACATACTGAAAAAACAAAATTAAAAAAAAGAG 171

Qy  18 GluLeuGluIleIleGlyGlySerGlySerLeuSerThrPhePheArgLeuPheAsn 37
         ||||||||||  ||::||  :::  |||||||||  :::  |||
Db 172 CATTTGCTACTATCATCCCAGAGTTCTCGGGATGTTCAACATTTCATCAATCAATCAAT 231

Qy  38 ArgSerPheThrGlnAlaLeu 44
         ::||| |||::::||
Db 232 CAATCAATCAATCAATCAATT 252
```

RESULT 15
AZ149972
LOCUS   AZ149972   483 bp   DNA   linear   GSS 28-AUG-2000
DEFINITION SP_0037_B1_E08_T7 Strongylocentrotus purpuratus, purple sea urchin,
            sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=37 Col=15 Row=J, DNA sequence.
ACCESSION   AZ149972
VERSION   AZ149972.1 GI:8301873
KEYWORDS   GSS.
SOURCE   Strongylocentrotus purpuratus.
ORGANISM   Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.

REFERENCE   1 (bases 1 to 483)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
            Hood,L.
  TITLE   A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT   Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 37 row: J column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 483.

```
FEATURES      Location/Qualifiers
     source   1..483
              /organism="Strongylocentrotus purpuratus"
              /db_xref="taxon:7668"
              /clone="plate=37 Col=15,Row=J"
              /clone_lib="Strongylocentrotus purpuratus, purple sea
              urchin, sperm genomic BAC library"
              /note="organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
              DH10B"
```

BASE COUNT   151 a   99 c   78 g   142 t
ORIGIN

Alignment Scores:
Pred. No.:   26.4   Length: 483
Score: 60.00   Matches: 14
Percent Similarity: 62.16%   Conservative: 9
Best Local Similarity: 37.84%   Mismatches: 12
Query Match: 26.43%   Indels: 2
DB: 17   Gaps: 1

US-09-833-017B-2 (1-46) x AZ149972 (1-483)

```
Qy   5 LeuSerLeuLysAsnAspPheLysGlyLeuLysThrAspGluGluGluIleIleIleGly 24
         |||||||:::||||||||::||||||  |||||  ||||||||||
Db 324 ATTGCTCTGAAGAACGATTACAGTGAATTATTTGAGATGAGAAGTAATNTCGTCATGGGA 383

Qy  25 GlySerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThr 41
         |||  :::  ||  |||  ||::  |||  |||::||
Db 384 GGGTGCTACAATGCGTCCCGTTTT------ATCTTCGGTCGAAATAGAACG 428
```

Search completed: November 8, 2002, 23:39:44
Job time : 1814.42 secs

```
                GenCore version 5.1.3
        Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    November  5, 2002, 10:55:02 ; Search time 37.0746 Seconds
                                (without alignments)
                        137.814 Million cell updates/sec

Title:          US-09-833-017B-2
Perfect score:  227
Sequence:     1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters:      747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  A_Geneseq_032802:*
     1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
     2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
     3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
     4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
     5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
     6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
     7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
     8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
     9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
    10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
    11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
    12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
    13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
    14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
    15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
    16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
    17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
    18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
    19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
    20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
    21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
    22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	58	25.6	173	18	AAW28307	Staphylococcus aur
2	56	24.7	939	22	AAU35204	Enterococcus faeca
3	55.5	24.4	553	8	AAP70176	Sequence of Newcas
4	55.5	24.4	553	15	AAR58858	Newcastle Disease
5	55.5	24.4	553	15	AAR49141	Newcastle disease
6	55.5	24.4	553	17	AAW06828	Newcastle disease
7	55.5	24.4	553	18	AAW10691	Newcastle disease
8	55.5	24.4	553	20	AAY21983	Seq ID No: 14 of U
9	55.5	24.4	553	21	AAB36039	Protein encoded by
10	55.5	24.4	553	21	AAY51231	Newcastle disease
11	55.5	24.4	553	21	AAY58183	NDV fusion (F) pro
12	55	24.2	583	22	AAG82132	S. epidermidis ope
13	54.5	24.0	508	21	AAG44399	Arabidopsis thalia
14	54.5	24.0	515	21	AAG44398	Arabidopsis thalia
15	54	23.8	96	14	AAR35757	Monellin protein.
16	54	23.8	97	14	AAR34192	Monellin protein.
17	54	23.8	2353	17	AAR99393	Haemophilus adhesi
18	54	23.8	2411	21	AAB23860	Haemophilus influe
19	53.5	23.6	553	12	AAR10065	Newcastle disease
20	53	23.3	96	21	AAG57201	Arabidopsis thalia
21	53	23.3	513	15	AAR54214	L.lactis branched
22	52	22.9	148	22	ABG28821	Novel human diagno
23	52	22.9	365	22	AAM78620	Human protein SEQ
24	52	22.9	381	22	AAM79604	Human protein SEQ
25	52	22.9	381	22	AAM79605	Human protein SEQ
26	52	22.9	382	22	AAM78621	Human protein SEQ
27	52	22.9	516	21	AAY52007	M. jannaschii MJ08
28	52	22.9	516	21	AAY51636	M. jannaschii MJ08
29	51.5	22.7	177	22	ABG17731	Novel human diagno
30	51.5	22.7	257	22	AAG98329	Escherichia coli p
31	51.5	22.7	306	21	AAB15918	E. coli proliferat
32	51.5	22.7	306	22	AAG98991	E. coli growth and
33	51.5	22.7	591	22	ABG10345	Novel human diagno
34	51.5	22.7	2737	22	ABG18288	Novel human diagno
35	51.5	22.7	3076	22	ABG07038	Novel human diagno
36	51	22.5	220	22	AAU37294	Staphylococcus aur
37	51	22.5	292	22	AAG74939	Human colon cancer
38	51	22.5	313	18	AAW21821	Mutant uracil DNA
39	51	22.5	313	18	AAW21822	Mutant uracil DNA
40	51	22.5	313	18	AAW21824	Mutant uracil DNA
41	51	22.5	313	18	AAW21825	Mutant uracil DNA
42	51	22.5	313	18	AAW21814	Uracil DNA glycosy
43	51	22.5	313	18	AAW21815	Cytosine DNA glyco
44	51	22.5	313	18	AAW21816	Thymine DNA glycos
45	51	22.5	313	18	AAW21819	Mutant uracil DNA

ALIGNMENTS

```
RESULT 1
AAW28307
ID   AAW28307 standard; Protein; 173 AA.
XX
AC   AAW28307;
XX
DT   01-SEP-1998  (first entry)
XX
DE   Staphylococcus aureus protein of unknown function.
XX
KW   Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW   Staphylococcal gene; regulatory element; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 123
FT                   /note= "X is not defined in the specification"
FT   Misc-difference 132
FT                   /note= "X is not defined in the specification"
XX
PN   WO9730070-A1.
XX
PD   21-AUG-1997.
XX
PF   19-FEB-1997;   97WO-US02318.
XX
PR   20-FEB-1996;   96US-0011888.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
```

```
PI  Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX
DR  WPI; 1997-42969/39.
DR  N-PSDB; AAT84204.
XX
PT  Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT  to isolate antimicrobial compounds, and in vaccines against S.
PT  aureus infection
XX
PS  Claim 6; Page 597-598; 989pp; English.
XX
CC  The present sequence represents a Staphylococcus aureus protein of
CC  unknown function. The DNA sequence was isolated from a library of
CC  clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
CC  be used in the construction of ribozymes and antisense sequences to
CC  control the expression of Staphylococcal genes. The DNA sequence is
CC  also useful as a source of regulatory elements for the control of
CC  bacterial gene expression. The present protein may be used to produce
CC  vaccines to enable a host to produce specific antibodies with
CC  antibacterial action. These vaccines and antibodies would protect
CC  a host against invasion by S. aureus, and conditions relating to
CC  Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC  skin syndrome, and toxic shock syndrome.
XX
SQ  Sequence  173 AA;

Query Match    25.6%;  Score 58;  DB 18;  Length 173;
Best Local Similarity  34.3%;  Pred. No. 3.3;
Matches  12;  Conservative  8;  Mismatches  15;  Indels  0;  Gaps  0;

Qy   9  NDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQA 43
          |||:||:| :::|:|:|:|:||:|
Db  86  NDFPELFTDTLVNSISANKDITKWFQTYNKSLLSA 120


RESULT 2
AAU35204
ID  AAU35204 standard; Protein; 939 AA.
XX
AC  AAU35204;
XX
DT  13-FEB-2002  (first entry)
XX
DE  Enterococcus faecalis cellular proliferation protein #491.
XX
KW  Antisense; prokaryotic cellular proliferation protein;
KW  antibiotic; antibacterial; drug design.
XX
OS  Enterococcus faecalis.
XX
PN  WO200170955-A2.
XX
PD  27-SEP-2001.
XX
PF  21-MAR-2001; 2001WO-US09180.
XX
PR  21-MAR-2000; 2000US-191078P.
PR  23-MAY-2000; 2000US-206848P.
PR  26-MAY-2000; 2000US-207727P.
PR  23-OCT-2000; 2000US-242578P.
PR  27-NOV-2000; 2000US-253625P.
PR  22-DEC-2000; 2000US-257931P.
PR  16-FEB-2001; 2001US-269308P.
XX
PA  (ELIT-) ELITRA PHARM INC.
XX
PI  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI  Yamamoto RT, Xu HH;
XX
DR  WPI; 2001-611495/70.
DR  N-PSDB; AAS53063.
XX
PT  New polynucleotides for the identification and development of
```

```
PT  antibiotics, comprise sequences of antisense nucleic acids -
XX
PS  Example 3; Seq ID No 10797; 51pp; English.
XX
CC  The invention relates to antisense inhibitors of genes essential to
CC  prokaryotic cellular proliferation, their use in identifying the
CC  genes, their use in the discovery of novel antibiotics, the essential
CC  genes themselves and the encoded proteins. The prokaryotes used are
CC  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC  invention is also useful for the identification of potential new targets
CC  for antibiotic development. The antisense nucleic acids can also be used
CC  to identify proteins used in proliferation, to express these proteins,
CC  and to obtain antibodies capable of binding to the expressed proteins.
CC  The proteins can be used to screen compounds in rational drug discovery
CC  programmes. The antisense nucleic acid sequence is also useful to screen
CC  for homologous nucleic acids which are required for cell proliferation in
CC  a wide variety of organisms. The present sequence represents an
CC  essential prokaryotic cellular proliferation protein.
CC  Note: The sequence data for this patent did not form part
CC  of the printed specification, but was obtained in electronic
CC  format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence  939 AA;

Query Match    24.7%;  Score 56;  DB 22;  Length 939;
Best Local Similarity  33.3%;  Pred. No. 45;
Matches  11;  Conservative  7;  Mismatches  15;  Indels  0;  Gaps  0;

Qy    5  LSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFN 37
           :::|:||:| |: |||: ||| :|| |
Db  330  IDMDTPFEELPADQQEIILNGSGEKNFHFHYEN 362


RESULT 3
AAP70176
ID  AAP70176 standard; protein; 553 AA.
XX
AC  AAP70176;
XX
DT  03-APR-1991  (first entry)
XX
DE  Sequence of Newcastle Disease Virus (NDV) F polypeptide.
XX
KW  Epitope; probe; diagnosis.
XX
OS  Newcastle Disease Virus.
XX
PN  EP227414-A.
XX
PD  01-JUL-1987.
XX
PF  16-DEC-1986;  86EP-0309804.
XX
PR  15-JUL-1986;  86US-0885765.
PR  18-DEC-1985;  85GB-0031147.
PR  14-APR-1986;  86GB-0009037.
XX
PA  (NATR ) NATIONAL RES DEV CORP.
XX
PI  Bingham RW, Chambers P, Emmerson PT, Millar NS;
XX
DR  WPI; 1987-179630/26.
DR  N-PSDB; AAN70261.
XX
PT  Newcastle disease virus gene clones - comprise polynucleotide(s)
PT  encoding the HN and/or F protein of Newcastle disease virus RNA
XX
PS  Example; pages 11-16; 22pp; English.
XX
CC  An artificial polynucleotide encoding an HN and/or F polypeptide of
CC  Newcastle Disease Virus (NDV) DNA, a bioprecursor of the
```

```
CC   polypeptide, or an epitopic portion of the polypeptide or an
CC   artificial nucleotide complementary to the polynucleotide are
CC   claimed. The polynucleotides are useful for preparing a probe for
CC   extracting similar genes from a gene library or for identifying the
CC   presence of NDV virons in a sample obtd. from poultry.
XX
SQ   Sequence    553 AA;

Query Match      24.4%;     Score 55.5;  DB 8;   Length 553;
Best Local Similarity   33.3%;    Pred. No. 29;
Matches   16;  Conservative    7;  Mismatches   20;  Indels    5;   Gaps    1;

Qy       4  TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK   46
             ||| | :||   :  |: :||| |: ::|| |: :|| :|||
Db     433  TLRLSGEFDATYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK  480

RESULT 4
AAR58858
ID   AAR58858 standard; Protein; 553 AA.
XX
AC   AAR58858;
XX
DT   13-APR-1995  (first entry)
XX
DE   Newcastle Disease Virus F gene product.
XX
KW   SfiI fragment; Newcastle Disease Virus; F gene.
XX
OS   Synthetic.
XX
PN   WO9419014-A.
XX
PD   01-SEP-1994.
XX
PF   28-FEB-1994;   94WO-US01826.
XX
PR   26-FEB-1993;   93US-0024156.
XX
PA   (JAPG ) NIPPON ZEON KK.
PA   (SYTR ) SYNTRO CORP.
XX
PI   Cochran MD;
XX
DR   WPI; 1994-294007/36.
DR   N-PSDB; AAQ68943.
XX
PT   New recombinant fowl pox virus for use in vaccines - contains
PT   genes expressing antigens of Newcastle disease virus and opt.
PT   infectious bronchitis virus
XX
PS   Disclosure; Page 70-72; 85pp; English.
XX
CC   Newcastle Disease Virus (NDV) HN and F genes were inserted as a
CC   SfiI fragment into the homology vector 443-88.8 at the unique
CC   SfiI site. The NDV HN and F genes were inserted in the same
CC   transcriptional orientation as the ORF in the parental homology
CC   vector. The sequence of SfiI fragment is in AAQ68943/R58858/AAR58859.
CC   The inserted SfiI fragment has the following structure:
CC   Junction A - Fragment 1 (HN, AAs 2-577) - Junction B - Fragment
CC   2 (F, AAs 1-553) - Junction C - Fragment 3 (pBR322) - Junction D.
CC   Fragment 1 is approx. 1811 bp AvaII to NaeI fragment of the NDV HN
CC   cDNA clone (B1 strain). Fragment 2 is an approx 1812 bp BamHI to
CC   PstI fragment of the full length NDV F cDNA (B1 strain). Fragment
CC   3 is an approx 235 bp PstI and ScaI fragment of the plasmid PBR322.
CC   The sequences of the Junctions are in AAQ68945, AAQ68946,
CC   AAQ68947,Q68948.
XX
SQ   Sequence    553 AA;

Query Match      24.4%;     Score 55.5;  DB 15;   Length 553;
Best Local Similarity   33.3%;    Pred. No. 29;
Matches   16;  Conservative    7;  Mismatches   20;  Indels    5;   Gaps    1;

Qy       4  TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK   46
             ||| | :||   :  |: :||| |: ::|| |: :|| :|||
Db     433  TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK  480

RESULT 5
AAR49141
ID   AAR49141 standard; Protein; 553 AA.
XX
AC   AAR49141;
XX
DT   14-APR-1995  (first entry)
XX
DE   Newcastle disease virus F gene product.
XX
KW   Newcastle disease virus; F gene; SfiI; vector 502-26.22.
XX
OS   Synthetic.
XX
PN   WO9419015-A.
XX
PD   01-SEP-1994.
XX
PF   28-FEB-1994;   94WO-US02252.
XX
PR   26-FEB-1993;   93US-0024156.
XX
PA   (SYTR ) SYNTRO CORP.
XX
PI   Cochran MD;
XX
DR   WPI; 1994-294008/36.
DR   N-PSDB; AAQ70570.
XX
PT   New recombinant fowl pox viruses - useful as vaccines against
PT   fowl pox virus, Newcastle Disease Virus and infectious
PT   laryngotracheitis virus.
XX
PS   Disclosure; Page 79-81; 97pp; English.
XX
CC   AAQ70570 is the SfiI fragment insert in Homology Vector 502-26.22
CC   contg. Newcastle disease virus (NDV) HN and F genes. The structure
CC   of the fragment is: 5'- Junction A - Fragment 1 (NDV HN AAs 2-577)
CC   - Junction B - Fragment 2 (NDV F AAs 1-553) - Junction C - Fragment
CC   3 (pBR322) - Junction D - 3'. Fragment 1 is AvaII to NaeI fragment
CC   of the full length NDV HN cDNA clone (B1 strain). Fragment 2 is
CC   BamHI to PstI fragment of the full length NDV F cDNA (B1 strain).
CC   Fragment 3 is a PstI and ScaI fragment of pBR322. The structures
CC   of the Junctions A, B, C and D are given in AAQ70552, AAQ70553,
CC   AAQ70554 and AAQ70555 respectively.
XX
SQ   Sequence    553 AA;

Query Match      24.4%;     Score 55.5;  DB 15;   Length 553;
Best Local Similarity   33.3%;    Pred. No. 29;
Matches   16;  Conservative    7;  Mismatches   20;  Indels    5;   Gaps    1;

Qy       4  TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK   46
             ||| | :||   :  |: :||| |: ::|| |: :|| :|||
Db     433  TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK  480

RESULT 6
AAW06828
ID   AAW06828 standard; Protein; 553 AA.
XX
AC   AAW06828;
XX
DT   04-MAR-1997  (first entry)
XX
DE   Newcastle disease virus F gene product.
XX
```

```
KW   Turkey herpes Virus; recombinant virus; vaccine; prophylaxis;
KW   immunisation; avian virus; infectious bronchitis virus;
KW   infectious bursal disease virus; Newcastle disease virus;
KW   Marek's disease virus; infectious laryngotracheitis virus; IBV;
KW   IBDV; NDV; MDV; ILV.
XX
OS   Newcastle disease virus.
XX
PN   WO9605291-A1.
XX
PD   22-FEB-1996.
XX
PF   09-AUG-1995;   95WO-US10245.
XX
PR   22-DEC-1994;   94US-0362240.
PR   09-AUG-1994;   94US-0288065.
XX
PA   (SYTR ) SYNTRO CORP.
XX
PI   Cochran MD, Junker DE,   Singer PA,   Wild MA;
XX
DR   WPI; 1996-139689/14.
DR   N-PSDB; AAT18203.
XX
PT   Recombinant turkey herpes viruses contg. foreign DNA encoding a
PT   cytokine - useful in vaccines to protect against Marek's disease
PT   virus and other avian viruses.
XX
PS   Disclosure; Page 179-180; 249pp; English.
XX
CC   Recombinant turkey herpes virus (rTH) which comprise a foreign DNA
CC   sequence encoding a cytokine inserted into a XhoI site within an
CC   EcoRI #9 genomic fragment, where the cytokine can be expressed in
CC   host cells infected with the virus can be used in vaccines to
CC   protect turkeys against avian viruses. The recombinant viruses can
CC   be used for immunising birds against infectious bronchitis virus
CC   (IBV), infectious bursal disease virus (IBDV), Marek's disease virus
CC   (MDV), infectious laryngotracheitis virus (ILV) and Newcastle disease
CC   virus (NDV). They may also be used in multivalent vaccines to
CC   protect against two or more of these avian viruses. This sequence
CC   is the product of the F gene of the Newcastle disease virus and is
CC   an antigen which can be used in the recombinant vaccines.
XX
SQ   Sequence   553 AA;
XX
Query Match      24.4%;   Score 55.5; DB 17;   Length 553;
Best Local Similarity   33.3%;   Pred. No. 29;
Matches  16; Conservative   7; Mismatches  20; Indels   5; Gaps   1;

Qy     4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||| :|       | |  :||  : ::||  :||     | | |
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 7
AAW10691
ID   AAW10691 standard; Protein; 553 AA.
XX
AC   AAW10691;
XX
DT   05-MAY-1997  (first entry)
XX
DE   Newcastle disease virus fusion protein.
XX
KW   Fowlpox virus; FPV; recombinant virus; vector; vaccine;
KW   immunisation; NDV; haemagglutinin; fusion protein; antigen;
KW   poultry.
XX
OS   Newcastle disease virus.
XX
PN   WO9640880-A1.
XX
PD   19-DEC-1996.
```

```
XX
PF   04-JUN-1996;   96WO-US11187.
XX
PR   07-JUN-1995;   95US-0484790.
XX
PA   (SYTR ) SYNTRO CORP.
XX
PI   Cochran MD, Junker DE,   Singer PA;
XX
DR   WPI; 1997-087060/08.
DR   N-PSDB; AAT48510.
XX
PT   New recombinant fowlpox virus - contg. a foreign DNA sequence
PT   inserted into the fowlpox virus genome, used for the prodn. of
PT   vaccines.
XX
PS   Disclosure; Page 109-110; 134pp; English.
XX
CC   Newcastle disease virus (NDV) haemagglutinin (HN) (AAW10690) and
CC   fusion (F) protein (AAW10691) are expressed by novel recombinant
CC   fowlpox virus (FPV). The genes (see also AAT48510) for HN and F can
CC   be inserted into homology vector 443-88.8 (see also AAT48511) at the
CC   unique SfiI site, yielding homology vector 502-26.22 (see also
CC   AAT48502-05), which can be used to insert the NDV HN and F genes
CC   into fowlpox virus.  The recombinant virus is used to deliver the
CC   vaccine antigens to poultry.
XX
SQ   Sequence   553 AA;
XX
Query Match      24.4%;   Score 55.5; DB 18;   Length 553;
Best Local Similarity   33.3%;   Pred No. 29;
Matches  16; Conservative   7; Mismatches  20; Indels   5; Gaps   1;

Qy     4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||| :|       | |  :||  : ::||  :||     | | |
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 8
AAY21983
ID   AAY21983 standard; Protein; 553 AA.
XX
AC   AAY21983;
XX
DT   07-SEP-1999  (first entry)
XX
DE   Seq ID No: 14 of US5925358.
XX
KW   Fowlpox virus; FPV; recombinant; vaccine; immunisation; chicken; NDV;
KW   Newcastle disease virus; NDV; Fowlpox; infectious Laryngotracheitis.
XX
OS   Fowlpox virus.
XX
PN   US5925358-A.
XX
PD   20-JUL-1999.
XX
PF   07-JUN-1995;   95US-0484575.
XX
PR   07-JUN-1995;   95US-0484575.
PR   26-FEB-1993;   93US-0024156.
PR   28-FEB-1994;   94WO-US02252.
XX
PA   (SYTR ) SYNTRO CORP.
XX
PI   Cochran MD, Junker DE;
XX
DR   WPI; 1999-418249/35.
DR   N-PSDB; AAX81147.
XX
PT   Fowlpox viruses, useful as vaccines for immunization of
PT   chickens/turkeys against Fowlpox and Newcastle disease virus
XX
```

PS   Disclosure; Column 61-70; 108pp; English.
XX
CC   The invention relates to a recombinant fowlpox virus (FPV) comprising
CC   a foreign DNA inserted into a region of the fowlpox virus genome
CC   corresponding to a 2.8 kb EcoRI fragment, capable of being expressed in a
CC   host cell. The virus is used as a vaccine for immunising chickens against
CC   Newcastle disease virus (NDV), Fowlpox, and infectious Laryngotracheitis.
XX
SQ   Sequence   553 AA;

Query Match   24.4%;   Score 55.5; DB 20; Length 553;
Best Local Similarity   33.3%;   Pred. No. 29;
Matches 16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| .|| :| : :||| : ::||: :|| ||:||| : |||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 9
AAB36039
ID   AAB36039 standard; Protein; 553 AA.
XX
AC   AAB36039;
XX
DT   02-MAR-2001 (first entry)
XX
DE   Protein encoded by NDV SfiI fragment.
XX
KW   Fowlpox virus; FPV; antiviral; antibacterial; vaccine;
KW   Newcastle's disease virus; NDV; Marek's disease;
KW   infectious laryngotracheitis.
XX
OS   Newcastle's disease virus.
XX
PN   US6136318-A.
XX
PD   24-OCT-2000.
XX
PF   07-JUN-1995;   95US-0486414.
XX
PR   26-FEB-1993;   93US-0024156.
PR   28-FEB-1994;   94WO-US02252.
XX
PA   (JUNK/) JUNKER D E.
PA   (COCH/) COCHRAN M D.
XX
PI   Cochran MD, Junker DE;
XX
DR   WPI; 2000-686071/67.
DR   N-PSDB; AAC67862.
XX
PT   New recombinant fowlpox virus useful as vaccines contains foreign DNA
PT   inserted into specific non-essential region of the genome
XX
PS   Disclosure; Column 71-74; 56pp; English.
XX
CC   The present sequence is provided in a specification relating to a
CC   recombinant fowlpox virus (FPV) that comprises a foreign DNA inserted
CC   within a 3.5 kb EcoRI fragment of the FPV genomic DNA. The foreign DNA
CC   can be expressed in host cells infected with FPV. The recombinant FPV
CC   may be used in vaccines to protect animals (especially chickens) against
CC   fowlpox and, depending on the source of the foreign DNA, other diseases,
CC   particularly Newcastle's disease, Marek's disease or infectious
CC   laryngotracheitis.
XX
SQ   Sequence   553 AA;

Query Match   24.4%;   Score 55.5; DB 21; Length 553;
Best Local Similarity   33.3%;   Pred. No. 29;
Matches 16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| .|| :| : :||| : ::||: :|| ||:||| : |||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 10
AAY51231
ID   AAY51231 standard; Protein; 553 AA.
XX
AC   AAY51231;
XX
DT   07-APR-2000 (first entry)
XX
DE   Newcastle disease virus LaSota genome encoded protein 4.
XX
KW   Avian-paramyxovirus; infection; lentogenic; F protein; vaccine;
KW   respiratory disease; gastrointestinal disease; poultry pathogen;
KW   local immunity.
XX
OS   Newcastle disease virus.
XX
PN   WO996045-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-NL00377.
XX
PR   19-JUN-1998;   98EP-0202054.
XX
PA   (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX
PI   Peeters BPH, De Leeuw OS, Koch G, Gielkens ALJ;
XX
DR   WPI; 2000-106102/09.
XX
PT   New avian paramyxovirus cDNA, useful for production of vaccine against
PT   Newcastle disease virus
XX
PS   Disclosure; Fig 3; 115pp; English.
XX
CC   This invention describes a novel avian-paramyxovirus cDNA (I) which
CC   comprises a nucleic acid sequence corresponding to the 5' terminal
CC   end of the genome of avian-paramyxovirus allowing the generation of
CC   an infectious copy of avian-paramyxovirus. The cell line is useful for
CC   the production of infectious lentogenic NDV. (Newcastle Disease virus)
CC   without the addition of exogenous proteolytic activity. Also it is
CC   possible to generate a stable transfected cell line that expresses the
CC   wild-type F protein in the virus envelope therefore providing infectious
CC   particles, useful in the form of a vaccine, especially against
CC   respiratory and/or gastrointestinal diseases. NDV can be easily cultured
CC   to very high titers in embryonated eggs. Mass culture of embryonated
CC   eggs is relatively cheap. NDV vaccines are relatively stable and can be
CC   simply administered by mass application methods e.g. drinking water or
CC   by spraying or by aerosol formation. The natural route of infection is
CC   by the respiratory and/or gastrointestinal tract which are also the major
CC   routes of infection of many other poultry pathogens. NDV can induce local
CC   immunity despite the presence of circulating maternal antibody. This
CC   sequence represents a protein encoded by the NDV strain LaSota genome
CC   which is described in the method of the invention.
XX
SQ   Sequence   553 AA;

Query Match   24.4%;   Score 55.5; DB 21; Length 553;
Best Local Similarity   33.3%;   Pred. No. 29;
Matches 16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| .|| :| : :||| : ::||: :|| ||:||| : |||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 11
AAY58183
ID   AAY58183 standard; Protein; 553 AA.

```
XX
AC   AAY58183;
XX
DT   14-MAR-2000  (first entry)
XX
DE   NDV fusion (F) protein.
XX
KW   Fowlpox virus; FPV; recombinant; antigenic protein; expression;
KW   infectious laryngotracheitis virus; ILTV; Newcastle disease virus; NDV;
KW   Marek's disease virus; cytokine; promoter; homologous recombination;
KW   homology vector; multivalent; live vaccine; fusion protein.
XX
OS   Newcastle disease virus.
XX
XX
FH   Key            Location/Qualifiers
FT   Domain         501..527
FT                  /note= "Transmembrane anchor domain"
XX
PN   US6001369-A.
XX
PD   14-DEC-1999.
XX
PF   07-JUN-1995;  95US-0477459.
XX
PR   26-FEB-1993;  93US-0024156.
PR   28-FEB-1994;  94WO-US02252.
XX
PA   (SYTR ) SYNTRO CORP.
XX
PI   Junker DE, Cochran MD;
XX
DR   WPI; 2000-071638/06.
DR   N-PSDB; AAZ49295.
XX
PT   Recombinant fowlpox virus useful as a vaccine for immunizing fowl
PT   against Marek's disease, Newcastle disease, Infectious
PT   Laryngotracheitis virus and/or fowlpox  -
XX
PS   Claim 5; Columns 71-74; 56pp; English.
XX
CC   The invention relates to a recombinant fowlpox virus (FPV)
CC   comprising a foreign DNA inserted into a 4.2 kb EcoRI fragment
CC   of the fowlpox virus genome. The foreign DNA is capable of being
CC   expressed in a host cell into which the fowlpox virus has been
CC   introduced and encodes an antigenic protein. The antigenic protein
CC   which may be expressed includes infectious laryngotracheitis virus
CC   (ILTV) glycoprotein B (gB) or glycoprotein D (gD, AAY58184), Newcastle
CC   disease virus (NDV) haemagglutinin (HN, AAY58182) or fusion (F) protein
CC   (AAY58183) and Marek's disease virus gB or gD. The foreign DNA may
CC   alternatively encode a cytokine such as chicken myelomonocytic growth
CC   factor (cMGF) or chicken interferon (cIFN). The foreign DNA in the
CC   recombinant FPV is under the control of one or more synthetic pox
CC   promoters, enabling control of strength and timing of heterologous
CC   gene expression. The synthetic pox virus promoters that may be used are
CC   based on promoters of the vaccinia virus and include early promoter 1
CC   (EP1), late promoter 1 (LP1), EP2 and LP2 (AAZ49291-Z49294,
CC   respectively). The recombinant FPV is generated via homologous
CC   recombination between FPV DNA and a homology vector containing the
CC   foreign DNA flanked by FPV sequences. The recombinant fowlpox viruses of
CC   the invention are used as multivalent live vaccines for immunising fowl
CC   against Marek's disease virus, NDV, ILTV and/or fowlpox virus. The
CC   present sequence represents NDV fusion (F) protein.
XX
SQ   Sequence  553 AA;
XX

Query Match             24.4%;  Score 55.5; DB 21; Length 553;
Best Local Similarity   33.3%;  Pred. No. 29;
Matches   16;  Conservative  7;  Mismatches   20;  Indels   5;  Gaps   1;

Qy     4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
           || :||  .-||..:||  :: ||||:.:||.| ||.||| |
Db   433 TLRLSGEFDVTYQKNISIQDSQVITGNLDISTELGNVNNSISNALNK 480
```

```
RESULT 12
AAG82132
ID   AAG82132 standard; Protein; 583 AA.
XX
AC   AAG82132;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis open reading frame protein sequence SEQ ID NO:1358.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW   vaccination; endocarditis.
XX
OS   Staphylococcus epidermidis.
XX
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000;  2000WO-US30782.
XX
PR   09-NOV-1999;  99US-0164258.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
DR   N-PSDB; AAH52982.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis -
XX
PS   Claim 18; Page 386; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC   (I) and (II) can have antibacterial activity and therefore can be used
CC   in vaccination. The nucleic acids (I) may be used to produce the
CC   S. epidermidis polypeptides (II) via the production of vectors
CC   containing them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to
CC   AAH55098 represent oligonucleotide sequences and primers which are used
CC   in the exemplification of the present invention.
CC   N.B. The present invention specifically claims all the polynucleotide
CC   sequences given in the sequence listing of the present specification,
CC   however the sequence listing only goes up to SEQ ID NO:4454 so even
CC   though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC   no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ   Sequence  583 AA;
XX

Query Match             24.2%;  Score 55; DB 22; Length 583;
Best Local Similarity   38.7%;  Pred. No. 36;
Matches   12;  Conservative  10;  Mismatches   5;  Indels   4;  Gaps   1;

Qy     1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLST 31
           :|.:|:    ||:::||:|| :|| | ::|| |
Db     8 IKETI----DFEQVETDQLFLISGKTGSGKT 34

RESULT 13
AAG44399
ID   AAG44399 standard; Protein; 508 AA.
XX
AC   AAG44399;
```

```
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 55612.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.          PR   18-JUN-1999; 99US-0139750.
PR   05-MAR-1999; 99US-0123180.          PR   18-JUN-1999; 99US-0139763.
PR   09-MAR-1999; 99US-0123548.          PR   21-JUN-1999; 99US-0139817.
PR   23-MAR-1999; 99US-0125788.          PR   22-JUN-1999; 99US-0139899.
PR   25-MAR-1999; 99US-0126264.          PR   23-JUN-1999; 99US-0140353.
PR   29-MAR-1999; 99US-0126785.          PR   23-JUN-1999; 99US-0140354.
PR   01-APR-1999; 99US-0127462.          PR   24-JUN-1999; 99US-0140695.
PR   06-APR-1999; 99US-0128234.          PR   28-JUN-1999; 99US-0140823.
PR   08-APR-1999; 99US-0128714.          PR   29-JUN-1999; 99US-0140991.
PR   16-APR-1999; 99US-0129845.          PR   30-JUN-1999; 99US-0141287.
PR   19-APR-1999; 99US-0130077.          PR   01-JUL-1999; 99US-0141842.
PR   19-APR-1999; 99US-0130449.          PR   01-JUL-1999; 99US-0142154.
PR   21-APR-1999; 99US-0130510.          PR   02-JUL-1999; 99US-0142055.
PR   23-APR-1999; 99US-0130891.          PR   06-JUL-1999; 99US-0142390.
PR   28-APR-1999; 99US-0131449.          PR   08-JUL-1999; 99US-0142803.
PR   30-APR-1999; 99US-0132048.          PR   09-JUL-1999; 99US-0142920.
PR   30-APR-1999; 99US-0132407.          PR   12-JUL-1999; 99US-0142977.
PR   04-MAY-1999; 99US-0132484.          PR   13-JUL-1999; 99US-0143542.
PR   05-MAY-1999; 99US-0132485.          PR   14-JUL-1999; 99US-0143624.
PR   06-MAY-1999; 99US-0132486.          PR   15-JUL-1999; 99US-0144005.
PR   06-MAY-1999; 99US-0132487.          PR   16-JUL-1999; 99US-0144085.
PR   07-MAY-1999; 99US-0132863.          PR   16-JUL-1999; 99US-0144086.
PR   11-MAY-1999; 99US-0134256.          PR   19-JUL-1999; 99US-0144325.
PR   14-MAY-1999; 99US-0134218.          PR   19-JUL-1999; 99US-0144331.
PR   14-MAY-1999; 99US-0134219.          PR   19-JUL-1999; 99US-0144332.
PR   14-MAY-1999; 99US-0134221.          PR   19-JUL-1999; 99US-0144333.
PR   18-MAY-1999; 99US-0134370.          PR   19-JUL-1999; 99US-0144334.
PR   19-MAY-1999; 99US-0134768.          PR   19-JUL-1999; 99US-0144335.
PR   19-MAY-1999; 99US-0134941.          PR   20-JUL-1999; 99US-0144352.
PR   20-MAY-1999; 99US-0135124.          PR   20-JUL-1999; 99US-0144632.
PR   21-MAY-1999; 99US-0135353.          PR   20-JUL-1999; 99US-0144884.
PR   24-MAY-1999; 99US-0135629.          PR   21-JUL-1999; 99US-0144814.
PR   25-MAY-1999; 99US-0136021.          PR   21-JUL-1999; 99US-0145086.
PR   27-MAY-1999; 99US-0136392.          PR   21-JUL-1999; 99US-0145088.
PR   28-MAY-1999; 99US-0136782.          PR   22-JUL-1999; 99US-0145085.
PR   01-JUN-1999; 99US-0137222.          PR   22-JUL-1999; 99US-0145087.
PR   03-JUN-1999; 99US-0137528.          PR   22-JUL-1999; 99US-0145089.
PR   04-JUN-1999; 99US-0137502.          PR   22-JUL-1999; 99US-0145192.
PR   07-JUN-1999; 99US-0137724.          PR   23-JUL-1999; 99US-0145145.
PR   08-JUN-1999; 99US-0138094.          PR   23-JUL-1999; 99US-0145218.
PR   10-JUN-1999; 99US-0138540.          PR   23-JUL-1999; 99US-0145224.
PR   10-JUN-1999; 99US-0138847.          PR   26-JUL-1999; 99US-0145276.
PR   14-JUN-1999; 99US-0139119.          PR   27-JUL-1999; 99US-0145913.
PR   16-JUN-1999; 99US-0139452.          PR   27-JUL-1999; 99US-0145918.
PR   16-JUN-1999; 99US-0139453.          PR   27-JUL-1999; 99US-0145919.
PR   17-JUN-1999; 99US-0139492.          PR   28-JUL-1999; 99US-0145951.
PR   18-JUN-1999; 99US-0139454.          PR   02-AUG-1999; 99US-0146386.
PR   18-JUN-1999; 99US-0139455.          PR   02-AUG-1999; 99US-0146388.
PR   18-JUN-1999; 99US-0139456.          PR   02-AUG-1999; 99US-0146389.
PR   18-JUN-1999; 99US-0139457.          PR   03-AUG-1999; 99US-0147038.
PR   18-JUN-1999; 99US-0139458.          PR   04-AUG-1999; 99US-0147204.
PR   18-JUN-1999; 99US-0139459.          PR   04-AUG-1999; 99US-0147302.
PR   18-JUN-1999; 99US-0139460.          PR   05-AUG-1999; 99US-0147192.
PR   18-JUN-1999; 99US-0139461.          PR   05-AUG-1999; 99US-0147260.
PR   18-JUN-1999; 99US-0139462.          PR   06-AUG-1999; 99US-0147303.
PR   18-JUN-1999; 99US-0139463.          PR   06-AUG-1999; 99US-0147416.
                                         PR   09-AUG-1999; 99US-0147493.
                                         PR   09-AUG-1999; 99US-0147935.
                                         PR   10-AUG-1999; 99US-0148171.
                                         PR   11-AUG-1999; 99US-0148319.
                                         PR   12-AUG-1999; 99US-0148341.
                                         PR   13-AUG-1999; 99US-0148565.
                                         PR   13-AUG-1999; 99US-0148684.
                                         PR   16-AUG-1999; 99US-0149368.
                                         PR   17-AUG-1999; 99US-0149175.
                                         PR   18-AUG-1999; 99US-0149426.
                                         PR   20-AUG-1999; 99US-0149722.
                                         PR   20-AUG-1999; 99US-0149723.
                                         PR   20-AUG-1999; 99US-0149929.
                                         PR   23-AUG-1999; 99US-0149902.
                                         PR   23-AUG-1999; 99US-0149930.
                                         PR   25-AUG-1999; 99US-0150566.
                                         PR   26-AUG-1999; 99US-0150884.
```

```
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   04-MAY-1999;   99US-0132407.
PR   05-MAY-1999;   99US-0132484.
PR   06-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   07-MAY-1999;   99US-0132487.
PR   11-MAY-1999;   99US-0132863.
PR   14-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   18-MAY-1999;   99US-0134370.
PR   19-MAY-1999;   99US-0134768.
PR   20-MAY-1999;   99US-0134941.
PR   21-MAY-1999;   99US-0135124.
PR   24-MAY-1999;   99US-0135353.
PR   25-MAY-1999;   99US-0135629.
PR   27-MAY-1999;   99US-0136021.
PR   28-MAY-1999;   99US-0136392.
PR   01-JUN-1999;   99US-0137782.
PR   03-JUN-1999;   99US-0137222.
PR   04-JUN-1999;   99US-0137528.
PR   07-JUN-1999;   99US-0137502.
PR   08-JUN-1999;   99US-0137724.
PR   10-JUN-1999;   99US-0138094.
PR   14-JUN-1999;   99US-0138540.
PR   16-JUN-1999;   99US-0138847.
PR   17-JUN-1999;   99US-0139119.
PR   18-JUN-1999;   99US-0139452.
PR   18-JUN-1999;   99US-0139453.
PR   18-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
```

```
Query Match        24.0%;      Score 54.5;   DB 21;   Length 508;
Best Local Similarity  39.6%;  Pred. No. 36;
Matches  21;  Conservative  8;  Mismatches  11;  Indels  13;  Gaps  5;

Qy   1 MKKTLSLKND-FKE-----IKTDELEIII-GGSGSLS---TF---FRLFNRSF 40
         !:!  :!!!!:!:  :!    !|!  :! !  !|:!:!
Db   1 MEKRSTLKNDSFVQEYNPVTETGSLSIIVLGASGDLAKKKTFPALFNLFHQGF 53
```

RESULT 14
AAG44398
ID   AAG44398 standard; Protein; 515 AA.
XX
AC   AAG44398;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 55611.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;

```
PR   28-JUN-1999;   99US-0140823.              PR   10-SEP-1999;   99US-0153070.
PR   29-JUN-1999;   99US-0140991.              PR   13-SEP-1999;   99US-0153758.
PR   30-JUN-1999;   99US-0141287.              PR   15-SEP-1999;   99US-0154018.
PR   01-JUL-1999;   99US-0141842.              PR   16-SEP-1999;   99US-0154039.
PR   01-JUL-1999;   99US-0142154.              PR   20-SEP-1999;   99US-0154779.
PR   02-JUL-1999;   99US-0142055.              PR   22-SEP-1999;   99US-0155139.
PR   06-JUL-1999;   99US-0142390.              PR   23-SEP-1999;   99US-0155486.
PR   08-JUL-1999;   99US-0142803.              PR   24-SEP-1999;   99US-0155659.
PR   09-JUL-1999;   99US-0142920.              PR   28-SEP-1999;   99US-0156458.
PR   12-JUL-1999;   99US-0142977.              PR   29-SEP-1999;   99US-0156596.
PR   13-JUL-1999;   99US-0143542.              PR   04-OCT-1999;   99US-0157117.
PR   14-JUL-1999;   99US-0143624.              PR   05-OCT-1999;   99US-0157753.
PR   15-JUL-1999;   99US-0144005.              PR   06-OCT-1999;   99US-0157865.
PR   16-JUL-1999;   99US-0144085.              PR   07-OCT-1999;   99US-0158029.
PR   16-JUL-1999;   99US-0144086.              PR   08-OCT-1999;   99US-0158232.
PR   19-JUL-1999;   99US-0144325.              PR   12-OCT-1999;   99US-0158369.
PR   19-JUL-1999;   99US-0144331.              PR   13-OCT-1999;   99US-0159293.
PR   19-JUL-1999;   99US-0144332.              PR   13-OCT-1999;   99US-0159294.
PR   19-JUL-1999;   99US-0144333.              PR   13-OCT-1999;   99US-0159295.
PR   19-JUL-1999;   99US-0144334.              PR   14-OCT-1999;   99US-0159329.
PR   20-JUL-1999;   99US-0144335.              PR   14-OCT-1999;   99US-0159330.
PR   20-JUL-1999;   99US-0144352.              PR   14-OCT-1999;   99US-0159331.
PR   20-JUL-1999;   99US-0144632.              PR   14-OCT-1999;   99US-0159637.
PR   21-JUL-1999;   99US-0144884.              PR   18-OCT-1999;   99US-0159638.
PR   21-JUL-1999;   99US-0144814.              PR   21-OCT-1999;   99US-0159584.
PR   21-JUL-1999;   99US-0145086.              PR   21-OCT-1999;   99US-0160741.
PR   22-JUL-1999;   99US-0145088.              PR   21-OCT-1999;   99US-0160767.
PR   22-JUL-1999;   99US-0145085.              PR   21-OCT-1999;   99US-0160768.
PR   22-JUL-1999;   99US-0145087.              PR   21-OCT-1999;   99US-0160770.
PR   22-JUL-1999;   99US-0145089.              PR   21-OCT-1999;   99US-0160814.
PR   23-JUL-1999;   99US-0145192.              PR   21-OCT-1999;   99US-0160815.
PR   23-JUL-1999;   99US-0145145.              PR   22-OCT-1999;   99US-0160980.
PR   23-JUL-1999;   99US-0145218.              PR   22-OCT-1999;   99US-0160981.
PR   23-JUL-1999;   99US-0145224.              PR   22-OCT-1999;   99US-0160989.
PR   26-JUL-1999;   99US-0145276.              PR   25-OCT-1999;   99US-0161404.
PR   27-JUL-1999;   99US-0145913.              PR   25-OCT-1999;   99US-0161405.
PR   27-JUL-1999;   99US-0145918.              PR   25-OCT-1999;   99US-0161406.
PR   27-JUL-1999;   99US-0145919.              PR   26-OCT-1999;   99US-0161359.
PR   28-JUL-1999;   99US-0145951.              PR   26-OCT-1999;   99US-0161360.
PR   02-AUG-1999;   99US-0146386.              PR   26-OCT-1999;   99US-0161361.
PR   02-AUG-1999;   99US-0146388.              PR   28-OCT-1999;   99US-0161920.
PR   02-AUG-1999;   99US-0146389.              PR   28-OCT-1999;   99US-0161992.
PR   03-AUG-1999;   99US-0147038.              PR   29-OCT-1999;   99US-0161993.
PR   04-AUG-1999;   99US-0147204.              PR   29-OCT-1999;   99US-0162142.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0147923.
PR   11-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   23-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150556.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
```

```
Query Match            24.0%;   Score 54.5;  DB 21;  Length 515;
Best Local Similarity  39.6%;   Pred. No. 36;
Matches  21; Conservative   8; Mismatches  11; Indels  13; Gaps   5;

Qy    1 MKKTLSLKND-FKE----IKTDELEIII-GGSGSLS---TF--FRLFNRSF  40
        |:| :|||| :.  :|  | |:| ||| |||:||| :. ||  ||::|
Db    8 MEKRSTLKNDSFVQEYNPVTFTGSLSIIVLGASGDLAKKTFPALFNLFHQGF  60


RESULT 15
AAR35757
ID   AAR35757 standard; Protein; 96 AA.
XX
AC   AAR35757;
XX
DT   29-JUL-1993  (first entry)
XX
DE   Monellin protein.
XX
KW   Sweet; heat; resistance; protease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 51
FT                   /label= GLY, SER, GLU
XX
PN   JP05070494-A.
```

XX
PD   23-MAR-1993.
XX
PF   30-MAY-1991;   91JP-0155713.
XX
PR   25-JUL-1990;   90JP-0196983.
XX
PA   (KIRI) KIRIN BREWERY KK.
XX
DR   WPI; 1993-131302/16.
XX
PT   Single-stranded monellin protein - has high heat resistance in
PT   cow pH range and high protease resistance
XX
PS   Claim 1; Page 2; 29pp; Japanese.
XX
CC   Monellin, is a sweet protein, having high thermal resistance at low
CC   pH's and also a high protease resistance.
XX
SQ   Sequence   96 AA;

Query Match   23.8%;   Score 54; DB 14; Length 96;
Best Local Similarity   41.4%;   Pred. No. 6;
Matches   12; Conservative   6; Mismatches   11; Indels   0; Gaps   0;

Qy   1 MKKTLSLKNDFKEIKTDELEIIIGGSGSL 29
            |||| :| |:|||| | :||:| | |
Db   42 MKKTIYEENXFREIKGYEYQLYVVASDKL 70

Search completed: November 5, 2002, 10:56:06
Job time : 40.0746 secs

```
                 GenCore version 5.1.3
          Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      November  5, 2002, 10:55:03 ; Search time 14.4179, Seconds
                                  (without alignments)
                                  77.929 Million cell updates/sec

Title:       US-09-833-017B-2
Perfect score: 227
Sequence:    1 MKKTLSLKNDFKEIKTDELE..........GSLSTPFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:      231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
             1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                         SUMMARIES
           %
Result     Query
No.  Score  Match Length DB ID                        Description
---- ----- ----- ------ -- ----------------------    ------------
  1  55.5  24.4  553   2  US-08-663-566A-13         Sequence 13, Appl
  2  55.5  24.4  553   2  US-08-484-575A-14         Sequence 14, Appl
  3  55.5  24.4  553   2  US-08-023-610-13          Sequence 13, Appl
  4  55.5  24.4  553   2  US-08-288-065A-13         Sequence 13, Appl
  5  55.5  24.4  553   2  US-08-362-240A-13         Sequence 13, Appl
  6  55.5  24.4  553   3  US-08-477-459-14          Sequence 14, Appl
  7  55.5  24.4  553   3  US-08-479-869-14          Sequence 14, Appl
  8  55.5  24.4  553   4  US-08-486-414-14          Sequence 14, Appl
  9  55.5  24.4  553   4  US-08-804-372A-11         Sequence 11, Appl
 10  55.5  24.4  553   5  PCT-US94-01826A-14        Sequence 14, Appl
 11  55.5  24.4  553   5  PCT-US94-02252A-14        Sequence 14, Appl
 12  55.5  24.4  553   5  PCT-US95-10245-13         Sequence 13, Appl
 13  55.5  24.4  553   6  5310678-1                 Patent No. 5310678
 14  55.5  24.4  837   4  US-09-564-805-228         Sequence 228, App
 15  54    23.8   97   4  US-09-242-690A-6          Sequence 6, Appli
 16  54    23.8  1912  1  US-08-409-995-4           Sequence 4, Appli
 17  54    23.8  1912  3  US-08-685-467-4           Sequence 4, Appli
 18  54    23.8  2353  4  US-09-377-155-33          Sequence 33, Appl
 19  54    23.8  2353  4  US-08-913-942-4           Sequence 4, Appli
 20  54    23.8  2353  4  US-09-669-974-33          Sequence 33, Appl
 21  54    23.8  2354  4  US-09-268-347-47          Sequence 47, Appl
 22  54    23.8  2411  4  US-09-268-347-36          Sequence 36, Appl
 23  53    23.3  513   1  US-08-403-866-1           Sequence 1, Appli
 24  51.5  22.7  244   4  US-08-919-573-2           Sequence 2, Appli
 25  51.5  22.7  244   4  US-08-919-573-4           Sequence 4, Appli
 26  49.5  21.8  436   3  US-08-486-099-94          Sequence 94, App
 27  49.5  21.8  436   3  US-08-360-107A-104        Sequence 104, App
 28  49.5  21.8  436   3  US-08-484-223B-94         Sequence 94, Appl
 29  49.5  21.8  436   3  US-08-919-597-94          Sequence 94, Appl
 30  49.5  21.8  436   3  US-08-475-668A-94         Sequence 94, Appl
 31  49.5  21.8  436   3  US-08-485-551A-94         Sequence 94, Appl
 32  49.5  21.8  436   3  US-08-471-913A-94         Sequence 94, Appl
 33  49.5  21.8  436   4  US-08-485-264A-94         Sequence 94, Appl
 34  49.5  21.8  202   4  US-08-474-349A-94         Sequence 94, App
 35  48.5  21.4  202   4  US-08-858-207A-332        Sequence 332, App
 36  48    21.1  720   3  US-09-257-799-48          Sequence 48, Appl
 37  48    21.1  720   2  US-08-920-919A-48         Sequence 48, Appl
 38  47.5  20.9   98   2  US-08-918-727-3           Sequence 3, Appli
 39  47.5  20.9   98   3  US-09-205-680A-3          Sequence 3, Appli
 40  47.5  20.9   98   3  US-09-048-889-11          Sequence 11, Appl
 41  47.5  20.9  270   2  US-08-852-743-5           Sequence 5, Appli
 42  47.5  20.9  270   2  US-09-185-370-5           Sequence 5, Appli
 43  47.5  20.9 15281  3  US-08-471-119A-2          Sequence 2, Appl
 44  47    20.7   92   3  US-08-813-884-53          Sequence 53, Appl
 45  47    20.7   93   3  US-08-813-884-52          Sequence 52, Appl

                         ALIGNMENTS

RESULT 1
US-08-663-566A-13
; Sequence 13, Application US/08663566A
; Patent No. 5853733
; GENERAL INFORMATION:
;  APPLICANT:  Cochran, Mark D
;  APPLICANT:  Macdonald, Richard D
;  TITLE OF INVENTION:  Recombinant Herpesvirus of Turkeys
;  TITLE OF INVENTION:  and Uses Thereof
;  NUMBER OF SEQUENCES:  56
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  John P. White
;   STREET:  1185 Avenue of the Americas
;   CITY:  New York
;   STATE:  New York
;   COUNTRY:  USA
;   ZIP:  10036
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  Patentin Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/663,566A
;   FILING DATE:  June 13, 1996
;   CLASSIFICATION:  435
;  ATTORNEY/AGENT INFORMATION:
;   NAME:  White, John P
;   REGISTRATION NUMBER:  28,678
;  TELECOMUNICATION INFORMATION:
;   TELEPHONE:  (212)278-0400
;   TELEFAX:  (212)391-0526
;   TELEX:  422523
;  INFORMATION FOR SEQ ID NO:  13:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH:  553 amino acids
;    TYPE:  amino acid
;    TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-663-566A-13

Query Match       24.4%;  Score 55.5;  DB 2;  Length 553;
Best Local Similarity 33.3%;  Pred. No. 3.4;
Matches  16; Conservative  7; Mismatches  20; Indels  5; Gaps  1;

Qy      4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTPFRLFNRSFTQALGK 46
          ||| :|    | ||: ::|| |:  ||| ::|| :|  ||:
Db    433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480
```

Wed Nov 13 14:19:06 2002

RESULT 2
US-08-484-575A-14
; Sequence 14, Application US/08484575A
; Patent No.5925358
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran and David E. Junker
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,575A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0450
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-484-575A-14

Query Match   24.4%;   Score 55.5; DB 2;   Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches 16;   Conservative 7;   Mismatches 20;   Indels 5;   Gaps 1;

Qy   4 TLSLKNDF----KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
          || :| || : :||: :| || :: :||
Db 433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 3
US-08-023-610-13
; Sequence 13, Application US/08023610
; Patent No.5928648
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D, Mark D
; APPLICANT: Macdonald Ph.D., Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
; TITLE OF INVENTION: and Uses Thereof
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/023,610
; FILING DATE: February 26, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-023-610-13

Query Match   24.4%;   Score 55.5; DB 2;   Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches 16;   Conservative 7;   Mismatches 20;   Indels 5;   Gaps 1;

Qy   4 TLSLKNDF----KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
          || :| || : :||: :| || :: :||
Db 433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 4
US-08-288-065A-13
; Sequence 13, Application US/08288065A
; Patent No.5961982
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Macdonald, Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-
; TITLE OF INVENTION: HVT-050 and Uses Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,065A
; FILING DATE: Aug-09-94
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0526
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-288-065A-13

Query Match   24.4%;   Score 55.5; DB 2;   Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches 16;   Conservative 7;   Mismatches 20;   Indels 5;   Gaps 1;

Qy   4 TLSLKNDF----KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
          || :| || : :||: :| || :: :||
Db 433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 5
US-08-362-240A-13

; Sequence 13, Application US/08362240A
; Patent No. 5965138
; GENERAL INFORMATION:
;   APPLICANT: Cochran, Mark D
;   APPLICANT: Junker, David
;   APPLICANT: Wild, Martha A
;   TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
;   NUMBER OF SEQUENCES: 60
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: John P. White
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/362,240A
;   FILING DATE: Dec-22-94
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White, John P
;   REGISTRATION NUMBER: 28,678
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212)278-0400
;   TELEFAX: (212)391-0526
;   TELEX: 422523
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 553 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-362-240A-13

Query Match   24.4%; Score 55.5; DB 2; Length 553;
Best Local Similarity   33.3%; Pred. No. 3.4;
Matches 16; Conservative   7; Mismatches 20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||| :| || : :||||| |: :|| | | :||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 6
US-08-477-459-14
; Sequence 14, Application US/08477459
; Patent No. 6001369
; GENERAL INFORMATION:
;   APPLICANT: Mark D. Cochran
;   TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: John P. White
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/477,459
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
;   NAME: White Esq, John P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 278-0400
;   TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 553 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-477-459-14

Query Match   24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity   33.3%; Pred. No. 3.4;
Matches 16; Conservative   7; Mismatches 20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||| :| || : :||||| |: :|| | | :||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 7
US-08-479-869-14
; Sequence 14, Application US/08479869
; Patent No. 6123949
; GENERAL INFORMATION:
;   APPLICANT: Cochran Ph.D., Mark D
;   TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and
;   TITLE OF INVENTION: Uses Thereof
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: John P. White
;   STREET: 30 Rockefeller Plaza
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10112
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/479,869
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/024,156
;   FILING DATE: 26-FEB-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White Esq, John P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212)977-9550
;   TELEFAX: (212)664-0525
;   TELEX: 422523
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 553 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-479-869-14

Query Match   24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity   33.3%; Pred. No. 3.4;
Matches 16; Conservative   7; Mismatches 20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||| :| || : :||||| |: :|| | | :||
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 8
US-08-486-414-14
; Sequence 14, Application US/08486414B
; Patent No. 6136318
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
; FILE REFERENCE: 42771D
; CURRENT APPLICATION NUMBER: US/08/486,414B
; CURRENT FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Fowlpox virus
US-08-486-414-14

Query Match   24.4%;   Score 55.5; DB 4; Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches   16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| .|| .:|| :|:||.:||| :|| .||| :|| | ||
Db 433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 9
US-08-804-372A-11
; Sequence 11, Application US/08804372A
; Patent No. 6183753
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Wild, Martha A.
; APPLICANT: Winslow, Barbara J.
; TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,372A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 2552/39115E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-804-372A-11

Query Match   24.4%;   Score 55.5; DB 4; Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches   16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

RESULT 10
PCT-US94-01826A-14
; Sequence 14, Application PC/TUS9401826A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox virus S-FPV-043 and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01826A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01826A-14

Query Match   24.4%;   Score 55.5; DB 5; Length 553;
Best Local Similarity   33.3%;   Pred. No. 3.4;
Matches   16; Conservative   7; Mismatches   20; Indels   5; Gaps   1;

Qy   4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| .|| .:|| :|:||.:||| :|| .||| :|| | ||
Db 433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 11
PCT-US94-02252A-14
; Sequence 14, Application PC/TUS9402252A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02252A

```
;    FILING DATE: 28-FEB-1994
;    CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;    NAME: White Esq, John P
;    TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (212)977-9550
;    TELEFAX:   (212)664-0525
;    TELEX:   422523
INFORMATION FOR SEQ ID NO: 14:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 553 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
PCT-US94-02252A-14
```

```
Query Match            24.4%;  Score 55.5;  DB 5;  Length 553;
Best Local Similarity  33.3%;  Pred. No. 3.4;
Matches  16;  Conservative  7;  Mismatches  20;  Indels  5;  Gaps  1;

Qy     4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
           ||| :|  |||: :| ::||  |||  : :||:  ||:|
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480
```

```
RESULT 12
PCT-US95-10245-13
; Sequence 13, Application PC/TUS9510245
; GENERAL INFORMATION:
; APPLICANT: SYNTRO CORPORATION
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10245
; FILING DATE: 09-AUG-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX:   (212)391-0526
; TELEX:   422523
INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-10245-13
```

```
Query Match            24.4%;  Score 55.5;  DB 5;  Length 553;
Best Local Similarity  33.3%;  Pred. No. 3.4;
Matches  16;  Conservative  7;  Mismatches  20;  Indels  5;  Gaps  1;

Qy     4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
           ||| :|  |||: :| ::||  |||  : :||:  ||:|
Db   433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 13
```

```
5310678-1
; Patent No. 5310678
; APPLICANT: Bingham, Richard W.;Chambers, Philip;Emmerson, Peter
; T.;Millar, Neil S.
; TITLE OF INVENTION: NEWCASTLE DISEASE VIRUS GENE CLONES
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/438,945
; FILING DATE: 17-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 885,765
; FILING DATE: 15-JUL-1986
; SEQ ID NO:1:
; LENGTH: 553
5310678-1
```

```
Query Match            24.4%;  Score 55.5;  DB 6;  Length 553;
Best Local Similarity  33.3%;  Pred. No. 3.4;
Matches  16;  Conservative  7;  Mismatches  20;  Indels  5;  Gaps  1;

Qy     4 TLSLKNDF-----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
           ||| :|  |||: :| ::||  |||  : :||:  ||:|
Db   433 TLRLSGEFDATYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480
```

```
RESULT 14
US-09-564-805-228
; Sequence 228, Application US/09564805
; Patent No. 6333403
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.F.
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
; TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes
; FILE REFERENCE: 2318-258
; CURRENT APPLICATION NUMBER: US/09/564,805
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/107,468
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: 09/434,382
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 228
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-564-805-228
```

```
Query Match            24.4%;  Score 55.5;  DB 4;  Length 837;
Best Local Similarity  34.1%;  Pred. No. 5.8;
Matches  14;  Conservative  9;  Mismatches  9;  Indels  9;  Gaps  1;

Qy     7 LKNDFKEIKTDELEIIIGGSGS-------LSTFFRLFNR 38
           |:|| :||: :|:|||| | |          ||:|||:|
Db   473 LPNCLEKIRRDDMEIVILGTGSSQPSKYRNVSAIFIDLFSR 513
```

```
RESULT 15
US-09-242-690A-6
; Sequence 6, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
; APPLICANT: KONDO, KEIJI
; APPLICANT: MIURA, YUTAKA
; TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
; TITLE OF INVENTION: SAME
; FILE REFERENCE: 04941/0118
; CURRENT APPLICATION NUMBER: US/09/242,690A
; CURRENT FILING DATE: 1999-02-23
```

```
;  PRIOR APPLICATION NUMBER: PCT/JP97/02924
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: JP 8/241062
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
;  SEQ ID NO 6
;    LENGTH: 97
;    TYPE: PRT
;    ORGANISM: Candida utilis
US-09-242-690A-6

Query Match      23.8%;  Score 54;  DB 4;  Length 97;
Best Local Similarity   41.4%;    Pred. No. 0.59;
Matches   12;  Conservative    6;  Mismatches   11;  Indels   0;  Gaps   0;

Qy    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSL 29
          |||:  .:|:||| |:.:. :  |
Db   43 MKKTIYEENGFREIKGYEYQLIVYASDKL 71

Search completed: November  5, 2002, 10:58:19
Job time : 16.4179 secs
```

```
                GenCore version 5.1.3
        Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      November  5, 2002, 10:55:02 ; Search time 17.8507 Seconds
                                           (without alignments)
                                           247.615 Million cell updates/sec

Title:         US-09-833-017B-2
Perfect score: 227
Sequence:      1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46
Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters:        283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_71:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	69	30.4	850	2	F95260	ABC transporter, p
2	69	30.4	850	2	A98126	conserved hypothet
3	60.5	26.7	475	2	S52893	hypothetical prot
4	58.5	25.8	41	2	H98126	competence stimula
5	58	25.6	1009	2	E89792	hypothetical prote
6	57.5	25.3	516	2	T52611	glucose-6-phosphat
7	56.5	24.9	398	2	AC3578	glycine betaine/l-
8	56.5	24.9	408	2	D86806	betaine ABC transp
9	56.5	24.9	593	2	C97848	ABC transporter AT
10	56.5	24.9	893	2	S64926	probable membrane
11	56	24.7	719	2	T12258	pyrroline-5-carbox
12	56	24.7	1250	2	E81339	probable restricti
13	55.5	24.4	416	2	AC1143	probable NAD(P)-de
14	55.5	24.4	416	2	AF1501	probable NAD(P)-de
15	55.5	24.4	553	1	VGNZNV	cell fusion glycop
16	55.5	24.4	553	2	VGNZGB	cell fusion glycop
17	55.5	24.4	553	2	D46329	cell fusion glycop
18	55.5	24.4	553	2	E46329	cell fusion glycop
19	55.5	24.4	553	2	G46329	cell fusion glycop
20	55.5	24.4	837	2	F96561	unknown protein [i
21	55	24.2	278	2	AB3091	hypothetical prote
22	55	24.2	278	2	H98195	hypothetical prote
23	55	24.2	608	2	S51790	centromere protein
24	54.5	24.0	230	2	D86746	cell-division ATP-
25	54.5	24.0	361	2	D98243	opuaa (AF234619) [
26	54.5	24.0	361	2	AH3042	hypothetical prote
27	54.5	24.0	373	2	C64419	histidinol-phospha
28	54.5	24.0	463	2	D97814	chromosomal replic
29	54.5	24.0	720	2	S73267	phenylalanine--tRN
30	54	23.8	226	1	C49769	gap junction prote
31	54	23.8	226	1	A33646	gap junction prote
32	54	23.8	243	2	C96946	glutamine ABC tran
33	54	23.8	708	2	T24727	hypothetical prote
34	54	23.8	1292	2	F64237	DNA-directed RNA p
35	53.5	23.6	165	2	S16722	coenzyme F420 hydr
36	53.5	23.6	241	2	AC1356	amino acid ABC tra
37	53.5	23.6	241	2	AD1726	amino acid ABC tra
38	53.5	23.6	242	2	AH1180	amino acid ABC tra
39	53.5	23.6	242	2	AI1537	amino acid ABC tra
40	53.5	23.6	243	2	D97008	ABC-type polar ami
41	53.5	23.6	246	2	G81409	probable phosphate
42	53.5	23.6	331	2	H81300	probable anion-upt
43	53.5	23.6	553	1	VGNZU1	cell fusion glycop
44	53.5	23.6	553	1	A46329	cell fusion glycop
45	53.5	23.6	553	1	B46329	cell fusion glycop

```
                              ALIGNMENTS

RESULT 1
F95260
ABC transporter, permease protein, probable SP2231 [imported] - Streptococcus pneumoniae
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: F95260
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: F95260
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-850 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK76279.1; PID:g14973742; GSPDB:GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP2231

Query Match              30.4%;     Score 69;   DB 2;   Length 850;
Best Local Similarity    33.3%;     Pred. No. 1.1;
Matches   19;  Conservative   10;  Mismatches   16;  Indels   12;  Gaps    2;

Qy       2 KKTLSLKNDF--------KEIKTDELIIIGGSGSLS---TFFRLFNRSFTQALGK 46
           |||::|  |||       | :|:|| :  :|:|:|:   ::||:|:|    ||:|
Db     692 KKTFTTDNVFSFFNLGYTKEKKTFNIVSFGNSQVSFESPTFYRLDTKTFTEAIQK 748

RESULT 2
A98126
conserved hypothetical protein spr2036 [imported] - Streptococcus pneumoniae (strain
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: A98126
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A98126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-850 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00838.1; PID:g15459743; GSPDB:GN00174
C;Genetics:
A;Gene: spr2036
```

Query Match   30.4%;   Score 69;   DB 2;   Length 850;
Best Local Similarity 33.3%;   Pred. No. 1.1;
Matches 19;   Conservative 10;   Mismatches 16;   Indels 12;   Gaps 2;

```
Qy    2 KKTLSLKNDF-------KEIKTDELEIIIGGSGSLS---TFFRLFNRSFTQALGK 46
         ||| | | |        |||  :: : ::| :: :|::|:|:|
Db  692 KKTFTTDNVFSFFNLGYTKEKKTNINVSFPGNSQVSFESPTFYRLDYKTFTEAIQK 748
```

RESULT 3
S52893
hypothetical protein YMR044w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9532.09
C;Species: Saccharomyces cerevisiae
C;Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C;Accession: S52893
R;Odell, C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A;Reference number: S52885
A;Accession: S52893
A;Molecule type: DNA
A;Residues: 1-475 <ODE>
A;Cross-references: EMBL:Z48502; NID:g695715; PIDN:CAA88410.1; PID:g695724; GSPDB:GN0001
C;Genetics:
A;Gene: MIPS:YMR044w
A;Map position: 13R

Query Match   26.7%;   Score 60.5;   DB 2;   Length 475;
Best Local Similarity 38.9%;   Pred. No. 6.9;
Matches 14;   Conservative 9;   Mismatches 12;   Indels 1;   Gaps 1;

```
Qy    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLF 36
         ||:|:|||  || ||| |||| |:  | |
Db  361 INQAVSIKDDF-EILLDELQIALDTRGSRNEFITIF 395
```

RESULT 4
H98126
competence stimulating peptide precursor (CSP) [imported] - Streptococcus pneumoniae (st
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: H98126
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: H98126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-41 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00845.1; PID:g15459751; GSPDB:GN00174
C;Genetics:
A;Gene: comC

Query Match   25.8%;   Score 58.5;   DB 2;   Length 41;
Best Local Similarity 44.4%;   Pred. No. 0.83;
Matches 16;   Conservative 5;   Mismatches 14;   Indels 1;   Gaps 1;

```
Qy    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLF 36
         ||||:||||  || ||| ||||||||||:||||
Db    1 MKNTVKLE-QFVALKEKDLQKIKGGEMRLSKFFRDF 35
```

RESULT 5
E89792
hypothetical protein SA0272 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Genetics:
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: E89792
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1009 <KUR>
A;Cross-references: GB:BA000018; PID:g13700198; PIDN:BAB41496.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0272

Query Match   25.6%;   Score 58;   DB 2;   Length 1009;
Best Local Similarity 34.3%;   Pred. No. 34;
Matches 12;   Conservative 8;   Mismatches 15;   Indels 0;   Gaps 0;

```
Qy    9 NDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQA 43
         ||| ||| |||| | :: :|: :|:|
Db  202 NDFPELFTDTLVNSISANKDITKWFQTYNKSLLSA 236
```

RESULT 6
T52611
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000
C;Accession: T52611
R;Wendt, U.K.; Hauschild, R.; Lange, C.; Pietersma, M.; Wenderoth, I.; von Schaewen,
Plant Mol. Biol. 40, 487-494, 1999
A;Title: Evidence for functional convergence of redox regulation in G6PDH isoforms of
A;Reference number: Z25284
A;Accession: T52611
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-516 <WEN>
A;Cross-references: EMBL:AJ010970; PIDN:CAB52674.1
C;Genetics:
A;Gene: acg9
C;Superfamily: glucose-6-phosphate dehydrogenase
C;Keywords: oxidoreductase

Query Match   25.3%;   Score 57.5;   DB 2;   Length 516;
Best Local Similarity 41.5%;   Pred. No. 19;
Matches 22;   Conservative 7;   Mismatches 11;   Indels 13;   Gaps 5;

```
Qy    1 MKKTLSLKND-FKE---IKTDELEIII-GGSGSLS---TF---FRLFNRSF 40
         |:| :||||  ||   |  | ||| ::|| || || |   :|      ||||||
Db    8 MEKRSTLKNDSFVKEYNPTTETGSLSIIVLGASGDLAKKKTFPALFNLFHQGF 60
```

RESULT 7
AC3578
glycine betaine/l-proline transport ATP-binding protein proV BMEII0548 [imported] - B
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AC3578
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3578
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53790.1; PID:g17984720; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0548
A;Map position: II

Query Match   24.9%;   Score 56.5;   DB 2;   Length 398;
Best Local Similarity   37.2%;   Pred. No. 19;
Matches 16;   Conservative   7;   Mismatches 19;   Indels   1;   Gaps   1;

Qy   4 TLSLKNDFKEIKTDELEIIIGGSGS-LSTFFRLFNRSFTQALG 45
             |:::| :|| |: :::|:|| ||| || ||||| | ||
Db   55 TIGVDNATFDIREGEVFVIMGLSGSGKSTLLRLLNRLIEPTAG 97

RESULT 8
D86806
betaine ABC transporter ATP binding protein busAA [imported] - Lactococcus lactis subsp.
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: D86806
R;Bolotin, A.; Winckler, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21351186; PMID:11337471
A;Accession: D86806
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-408 <STO>
A;Cross-references: GB:AE005176; PID:g12724443; PIDN:AAK05550.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: busAA

Query Match   24.9%;   Score 56.5;   DB 2;   Length 408;
Best Local Similarity   45.7%;   Pred. No. 19;
Matches 16;   Conservative   3;   Mismatches   5;   Indels   1;   Gaps   1;

Qy   13 EIKTDELEIIIGGSGS-LSTFFRLFNRSFTQALGK 46
             || |:::|| || :| ||| || || ||| |
Db   50 EINEGEIFVIMGLSGSGKSTLLRLLNRLIEPTSGK 84

RESULT 9
C97848
ABC transporter ATP-binding protein uup [imported] - Rickettsia conorii (strain Malish 7
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C;Accession: C97848
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID:21442074; PMID:11557893
A;Accession: C97848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-593 <KUR>
A;Cross-references: GB:AE006914; PIDN:AAL03725.1; PID:g15620316; GSPDB:GN00173
C;Genetics:
A;Gene: uup
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match   24.9%;   Score 56.5;   DB 2;   Length 593;
Best Local Similarity   37.0%;   Pred. No. 29;
Matches 17;   Conservative   5;   Mismatches 23;   Indels   1;   Gaps   1;

Qy   2 KKTLSLKNDFKEIKTDELEIIIGGSGS-LSTFFRLFNRSFTQALGK 46
             | | :: :| ||| :|:|| || :| | ||||| | ||
Db   298 KNTKIINNFSFVKNGEKIGIGANGSGKSTFIKLLTKQLTPESGK 343

RESULT 10
S64926
probable membrane protein YLR092w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein L2528
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 06-Feb-1998
C;Accession: S64926
R;Benes, V.; Rechmann, S.; Nentwich, U.; Schwager, C.; Ansorge, W.; Voss, H.

submitted to the Protein Sequence Database, May 1996
A;Reference number: S64920
A;Accession: S64926
A;Molecule type: DNA
A;Residues: 1-893 <BEN>
A;Cross-references: EMBL:Z73264; NID:g1360466; PID:e245799; PID:g1360467; MIPS:YLR092
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SUL2
A;Cross-references: SGD:S0004082; MIPS:YLR092w
A;Map position: 12R
C;Keywords: transmembrane protein
F;136-152/Domain: transmembrane #status predicted <TM1>
F;223-239/Domain: transmembrane #status predicted <TM2>
F;244-260/Domain: transmembrane #status predicted <TM3>
F;302-318/Domain: transmembrane #status predicted <TM4>
F;358-374/Domain: transmembrane #status predicted <TM5>
F;411-427/Domain: transmembrane #status predicted <TM6>
F;488-504/Domain: transmembrane #status predicted <TM7>
F;538-554/Domain: transmembrane #status predicted <TM8>
F;557-573/Domain: transmembrane #status predicted <TM9>

Query Match   24.9%;   Score 56.5;   DB 2;   Length 893;
Best Local Similarity   36.7%;   Pred. No. 46;
Matches 18;   Conservative   6;   Mismatches 18;   Indels   7;   Gaps   2;

Qy   1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRL----FNRSFTQA 43
             : :| ||| || :| || |||||||| : | :||:|
Db   426 ISKSFGRINDYKVVPDQEL-IAIGVSNLLGTFFNAYPATGSFSRSALKA 473

RESULT 11
T12258
pyrroline-5-carboxylate synthetase (EC 1.5.1.-) - common ice plant
C;Species: Mesembryanthemum crystallinum (common ice plant)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C;Accession: T12258
R;Michalowski, C.B.; Quigley-Landreau, F.; Bohnett, H.J.
submitted to the EMBL Data Library, May 1998
A;Description: Mesembryanthemum crystallinum pyrroline-5-carboxylate synthetase mRNA.
A;Reference number: 217475
A;Accession: T12258
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-719 <MIC>
A;Cross-references: EMBL:AF067967; NID:g3176964; PID:g3176965
C;Genetics:
A;Gene: P5CS
C;Superfamily: glutamate-5-semialdehyde dehydrogenase
C;Keywords: oxidoreductase

Query Match   24.7%;   Score 56;   DB 2;   Length 719;
Best Local Similarity   36.6%;   Pred. No. 42;
Matches 15;   Conservative   7;   Mismatches 13;   Indels   6;   Gaps   1;

Qy   6 SLKNDFKEIKTDELEIIIGGSGSLST------FFRLFNRSF 40
             :: || |: :|| || |||||||| :|| : ||||| |
Db   37 SLCEQLKELNSDGYEVILVTSGAVSAGRQRLRFRKLVNSSF 77

RESULT 12
E81339
probable restriction /modification enzyme Cj0690c [imported] - Campylobacter jejuni (
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: E81339
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Bashan, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: E81339
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-1250 <PAR>
A;Cross-references: GB:AL139076; GB:AL133076; GB:AL111168; NID:g6968128; PIDN:CAB72964.1; PID:g696814
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0690c

Query Match   24.7%;   Score 56;   DB 2;   Length 1250;
Best Local Similarity   40.0%;   Pred. No. 77;
Matches   14;   Conservative   4;   Mismatches   15;   Indels   2;   Gaps   1;

Qy   3 KTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFN 37
      | |||||:   : || ||| : : : ||
Db   380 KNLSFKRDYKTLSVEHLGTIY--EGLLSYFFEIAN 412

RESULT 13
AC1143
probable NAD(P)-dependent oxidoreductase homolog lmo0546 [imported] - Listeria monocytog
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AC1143
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1143
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-416 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98625.1; PID:g16409922; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0546

Query Match   24.4%;   Score 55.5;   DB 2;   Length 416;
Best Local Similarity   44.4%;   Pred. No. 26;
Matches   12;   Conservative   7;   Mismatches   7;   Indels   1;   Gaps   1;

Qy   2 KKTLSLKNDFKE-IKTDELEIIIGGSG 27
      :: | ||||| :| :|:|:|: :|
Db   68 KEKILLSNDFKEIIHSDLVEVIVDATG 94

RESULT 14
AF1501
probable NAD(P)-dependent oxidoreductase homolog lin0550 [imported] - Listeria innocua
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AF1501
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1501
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-416 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95782.1; PID:g16412990; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0550

Query Match   24.4%;   Score 55.5;   DB 2;   Length 416;
Best Local Similarity   44.4%;   Pred. No. 26;

Matches   12;   Conservative   7;   Mismatches   7;   Indels   7;   Gaps   1;

Qy   2 KKTLSLKNDFKE-IKTDELEIIIGGSG 27
      |:: | ||||| |: :|:|:|: :|
Db   68 KEKILLSNDFKEIIHSDLVEVIVDATG 94

RESULT 15
VGNZNV
cell fusion glycoprotein precursor - Newcastle disease virus (strains Beaudette C and
N;Contains: fusion glycoprotein F1; fusion glycoprotein F2
C;Species: Newcastle disease virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C;Accession: A27008; F46329
R;Chambers, P.; Millar, N.S.; Emmerson, P.T.
J. Gen. Virol. 67, 2685-2694, 1986
A;Title: Nucleotide sequence of the gene encoding the fusion glycoprotein of Newcastl
A;Reference number: A27008; MUID:87085486
A;Accession: A27008
A;Molecule type: mRNA
A;Residues: 1-553 <CHA>
A;Cross-references: GB:X04719; NID:g60933; PIDN:CAA28426.1; PID:g60934
A;Experimental source: strain Beaudette C
R;Toyoda, T.; Sakaguchi, T.; Hirota, H.; Gotoh, B.; Kuma, K.; Miyata, T.; Nagai, Y.
Virology 169, 273-282, 1989
A;Title: Newcastle disease virus evolution. II. Lack of gene recombination in generat
A;Reference number: A46329; MUID:89204898
A;Accession: F46329
A;Molecule type: genomic RNA
A;Residues: 1-553 <TOY>
A;Cross-references: GB:M24697; NID:g293929; PIDN:AAA46648.1; PID:g293930
A;Experimental source: strain BEA/45
C;Genetics:
A;Gene: F
C;Superfamily: parainfluenza virus cell fusion protein
C;Keywords: glycoprotein; membrane fusion; transmembrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-117/Product: cell fusion glycoprotein F2 #status predicted <FF2>
F;118-553/Product: cell fusion glycoprotein F1 #status predicted <FF1>
F;495-528/Domain: transmembrane #status predicted <TMN>
F;85,191,366,447,471/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   24.4%;   Score 55.5;   DB 1;   Length 553;
Best Local Similarity   33.3%;   Pred. No. 36;
Matches   16;   Conservative   7;   Mismatches   16;   Indels   20;   Gaps   1;

Qy   4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
      || |: || | : :|| | :|| : ||
Db   433 TLRLSGEFDATYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

Search completed: November   5, 2002, 10:56:40
Job time : 21.8507 secs

```
          GenCore version 5.1.3
      Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:  November 5, 2002, 10:55:03  ; Search time 10.9851 Seconds
                                        (without alignments)
                               162.138 Million cell updates/sec

Title:         US-09-833-017B-2
Perfect score: 227
Sequence:      1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:    105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : SwissProt_40:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60.5	26.7	475	1	YMT4_YEAST	Q04213 saccharomyc
2	58.5	25.8	41	1	CSP1_STRPN	Q54712 streptococc
3	56.5	24.9	893	1	SU12_YEAST	Q12325 saccharomyc
4	56	24.7	719	1	P5CS_MESCR	O65361 m delta 1-p
5	55.5	24.4	41	1	CSP_STROR	O33689 streptococc
6	55.5	24.4	553	1	VGLF_NDVB	P06156 newcastle d
7	55.5	24.4	553	1	VGLF_NDVH4	P33613 newcastle d
8	55.5	24.4	553	1	VGLF_NDVL	P33614 newcastle d
9	55.5	24.4	553	1	VGLF_NDVTG	P12571 newcastle d
10	55	24.2	608	1	CB32_YEAST	P40969 saccharomyc
11	54.5	24.0	373	1	HIS8_METJA	Q58365 methanococc
12	54.5	24.0	463	1	DNAA_RICCN	Q92b56 rickettsia
13	54.5	24.0	720	1	SYFB_PORPU	P51346 porphyra pu
14	54	23.8	226	1	CXB2_MOUSE	Q00977 mus musculu
15	54	23.8	226	1	CXB2_RAT	P21994 rattus norv
16	54	23.8	708	1	YN2B_CAEEL	P45972 caenorhabdi
17	54	23.8	1292	1	RPOC_MYCGE	P47582 mycoplasma
18	53.5	23.6	259	1	PSTB_EDWTA	Q9aml4 edwardsiell
19	53.5	23.6	553	1	VGLF_NDVD	P35936 newcastle d
20	53.5	23.6	553	1	VGLF_NDVQ	P33615 newcastle d
21	53.5	23.6	553	1	VGLF_NDVU	P12570 newcastle d
22	53.5	23.6	617	1	CYG2_HUMAN	O75343 homo sapien
23	53	23.3	274	1	PNK_AQUAE	O67055 aquifex aeo
24	53	23.3	400	1	PROV_ECOLI	P14175 escherichia
25	53	23.3	400	1	PROV_SALTY	P17328 salmonella
26	53	23.3	513	1	LEU1_LACLA	Q02141 lactococcus
27	53	23.3	591	1	YN48_YEAST	P42846 saccharomyc
28	53	23.3	1132	1	DNB1_HSV6U	P52338 human herpe
29	53	23.3	1132	1	DNB1_HSV6Z	P52538 human herpe
30	52.5	23.1	310	1	GTRB_BPP22	P57022 bacteriopha
31	52.5	23.1	564	1	Y61A_MYCPN	P75041 mycoplasma
32	51.5	22.7	41	1	CSP2_STROR	O33690 streptococc
33	51.5	22.7	240	1	PYRH_METJA	Q58656 methanococc
34	51.5	22.7	252	1	Y830_METJA	Q58240 methanococc
35	51.5	22.7	257	1	PSTB_ECOLI	P07655 escherichia
36	51.5	22.7	259	1	DNAA_MYCBO	P49991 mycobacteri
37	51.5	22.7	306	1	GTRB_ECOLI	P77293 escherichia
38	51	22.5	230	1	YP2A_STRAU	P03866 staphylococ
39	51	22.5	295	1	UNG_MOUSE	P97931 mus musculu
40	51	22.5	304	1	UNG_HUMAN	P13051 homo sapien
41	51	22.5	360	1	DFRA_DIACA	P51104 dianthus ca
42	51	22.5	397	1	Y636_METJA	Q58053 methanococc
43	51	22.5	461	1	Y061_MYCGE	P47307 mycoplasma
44	51	22.5	465	1	CG12_CANAL	P43062 candida alb
45	51	22.5	511	1	G6PD_EMENI	P41764 emericella

ALIGNMENTS

```
RESULT 1
YMT4_YEAST
ID   YMT4_YEAST    STANDARD;      PRT;    475 AA.
AC   Q04213;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 55.4 kDa protein in MCM1-NUP116 intergenic region.
GN   YMR044W OR YM9532.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / A8972;
RA   Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!-
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z48502; CAA88410.1; -.
DR   SGD; S0004647; YMR044W.
DR   InterPro; IPR000313; PWWP.
DR   Pfam; PF00855; PWWP; 1.
DR   SMART; SM00293; PWWP; 1.
KW   Hypothetical protein.
FT   DOMAIN     74     77       POLY-GLU.
FT   DOMAIN    186    195       POLY-GLU.
FT   DOMAIN    243    251       POLY-GLU.
SQ   SEQUENCE   475 AA;  55426 MW;  DFD6189E8117B060 CRC64;

Query Match         26.7%;  Score 60.5;  DB 1;  Length 475;
Best Local Similarity  38.9%;  Pred. No. 2.8;
Matches  14;  Conservative  9;  Mismatches  12;  Indels 1;  Gaps 1;

Qy    1 MKKTLSLKNDFKEIKTDELEIIGGSGSLSTFFRLF 36
          : :|:||:|| ||||:|: : :| : :|
Db  361 INQAVSIKDDF-EIILDELQIALDTRGSRNEFITIF 395

RESULT 2
CSP1_STRPN
ID   CSP1_STRPN    STANDARD;      PRT;     41 AA.
AC   Q54712;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Competence stimulating peptide type 1 precursor (CSP-1).
GN   COMC1 OR COMC.
```

```
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
RX   STRAIN=RX / CP1200;
RX   MEDLINE=9074663; PubMed=7479953;
RA   Haevarstein L.S., Coomaraswamy G., Morrison D.A.;
RT   "An unmodified heptadecapeptide pheromone induces competence for
RT   genetic transformation in Streptococcus pneumoniae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:11140-11144(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R6x;
RX   MEDLINE=97206147; PubMed=9157240;
RA   Cheng Q., Campbell E.A., Naughton A.M., Johnson S., Masure H.R.;
RT   "The com locus controls genetic transformation in Streptococcus
RT   pneumoniae.";
RL   Mol. Microbiol. 23:683-692(1997).
CC   -!- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
CC       FOR GENETIC TRANSFORMATION.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE COMC FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U33315; AAC44440.1; -.
DR   EMBL; U76218; AAC44895.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
KW   Pheromone; Competence.
FT   PROPEP        1     24
FT   CHAIN        25     41       COMPETENCE STIMULATING PEPTIDE TYPE 1.
SQ   SEQUENCE   41 AA;   4971 MW;  1448B414E980E86A CRC64;

Query Match             25.8%;  Score 58.5; DB 1; Length 41;
Best Local Similarity   44.4%;  Pred. No. 0.34;
Matches   16;  Conservative    5;  Mismatches   14;  Indels    1;  Gaps    1;

Qy   1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLF 36
        || |:|: :| :  :| ||| :|| ||||||
Db   1 MKNTVKLE-QFVALKEKDLQKIKGGEMRLSKFFRDF 35

RESULT 3
ID   SUL2_YEAST    STANDARD;    PRT;   893 AA.
AC   Q12325;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Sulfate permease 2 (High-affinity sulfate transporter 2).
GN   SUL2 OR SEL2 OR YLR092W OR L9449.1.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA   Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA   Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA   Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA   Taich A., Trevaskis E., Vignali D., Wilcox L., Wohldman P., Vaudin M.,
RA   Wilson R., Waterston R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benes V., Rechmann S., Nentwich U., Schwager C., Ansorge W., Voss H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=97207837; PubMed=9055073;
RA   Cherest H., Davidian J.C., Thomas D., Benes V., Ansorge W.,
RA   Surdin-Kerjan Y.;
RT   "Molecular characterization of two high affinity sulfate transporters
RT   in Saccharomyces cerevisiae.";
RL   Genetics 145:627-635(1997).
CC   -!- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U53880; AAB67596.1; -.
DR   EMBL; Z73364; CAA97653.1; -.
DR   EMBL; Z73265; CAA97655.1; -.
DR   EMBL; U53876; AAB67550.1; -.
DR   SGD; S0004082; SUL2.
DR   InterPro; IPR002645; STAS.
DR   InterPro; IPR001902; Sulfate_transp.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   PROSITE; PS01130; SULFATE_TRANSP; 1.
KW   Transport; Transmembrane; Glycoprotein.
FT   TRANSMEM    132    152       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
FT   TRANSMEM    189    209       POTENTIAL.
FT   TRANSMEM    222    242       POTENTIAL.
FT   TRANSMEM    245    265       POTENTIAL.
FT   TRANSMEM    306    326       POTENTIAL.
FT   TRANSMEM    351    371       POTENTIAL.
FT   TRANSMEM    400    420       POTENTIAL.
FT   TRANSMEM    444    464       POTENTIAL.
FT   TRANSMEM    484    504       POTENTIAL.
FT   TRANSMEM    539    559       POTENTIAL.
FT   TRANSMEM    560    580       POTENTIAL.
SQ   SEQUENCE  893 AA;  99650 MW;  67826955AC7C0BF5 CRC64;

Query Match             24.9%;  Score 56.5;  DB 1; Length 893;
Best Local Similarity   36.7%;  Pred. No. 19;
Matches   18;  Conservative    6;  Mismatches   18;  Indels    7;  Gaps    2;

Qy    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRL------FNRSFTQA 43
         :|: ||:|: |||| :|| :| :||| | :|| ::
Db  426 ISKSFGRINDYKVVPDQEL-IATGVSNLLGTFNAYPATGSFSRSALKA 473

RESULT 4
ID   P5CS_MESCR    STANDARD;    PRT;   719 AA.
AC   O65361;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Delta 1-Pyrroline-5-carboxylate synthetase (P5CS) [Includes: Glutamate
DE   5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK); Gamma-glutamyl
DE   phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde
DE   dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase)].
GN   P5CS.
OS   Mesembryanthemum crystallinum (Common ice plant).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
OX   NCBI_TaxID=3544;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Michalowski C.B., Quigley-Landreau F., Bohnert H.J.;
RT   "Mesembryanthemum crystallinum pyrroline-5-carboxylate synthetase
RT   mRNA.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: P5CS PLAYS A KEY ROLE IN PROLINE BIOSYNTHESIS, LEADING
CC       TO OSMOREGULATION IN PLANTS.
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
CC       phosphate
CC   -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC       NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC   -!- ENZYME REGULATION: FEEDBACK REGULATED BY PROLINE.
CC   -!- PATHWAY: FIRST AND SECOND STEPS IN PROLINE BIOSYNTHESIS PATHWAY.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN LEAVES AND IS
CC       INDUCIBLE IN ROOTS SUBJECTED TO SALT STRESS.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUTAMATE 5-
CC       KINASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GAMMA-
CC       GLUTAMYL PHOSPHATE REDUCTASE FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AF067967; AAC18862.1; -.
DR   Mendel; 29718; Mescr;1229;29718.
DR   InterPro; IPR001048; Aakinase.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR000965; GPR.
DR   InterPro; IPR001057; Glut_5_kinase.
DR   Pfam; PF00696; aakinase; 1.
DR   Pfam; PF00171; aldedh; 1.
DR   PRINTS; PR00474; GLU5KINASE.
DR   PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
DR   PROSITE; PS01223; PROA; 1.
KW   Proline biosynthesis; Multifunctional enzyme; Oxidoreductase; NADP;
KW   Transferase; Kinase.
FT   DOMAIN      1    293      GLUTAMATE 5-KINASE.
FT   DOMAIN    294    719      GAMMA-GLUTAMYL PHOSPHATE REDUCTASE.
SQ   SEQUENCE   719 AA;  77855 MW;  1510D11AF5559961 CRC64;

Query Match             24.7%;  Score 56;  DB 1;  Length 719;
Best Local Similarity   36.6%;  Pred. No. 17;
Matches   15;  Conservative   7;  Mismatches   13;  Indels   6;  Gaps   1;

Qy    6 SLKNDFKEIKTDELEIIIGGSGSLST------FFRLFNRSF 40
          ||  | |:|  :| ||:::| ||::|        :|:|| ||
Db   37 SLCEQLKELNSDGYEVILVTSGAVSAGRQRLRFRKLVNSSF 77


RESULT 5
CSP_STROR
ID   CSP_STROR     STANDARD;      PRT;    41 AA.
AC   O33689;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Competence stimulating peptide precursor (CSP).
GN   COMC.
OS   Streptococcus oralis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1303;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11427;
RX   MEDLINE=98012953; PubMed=9352904;
RA   Haevarstein L.S., Hakenbeck R., Gaustad P.;
RT   "Natural competence in the genus Streptococcus: evidence that
RT   streptococci can change phenotype by interspecies recombinational
RT   exchanges.";
RL   J. Bacteriol. 179:6589-6594(1997).
CC   -!- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
CC       FOR GENETIC TRANSFORMATION.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE COMC FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AJ000873; CAA04363.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
KW   Pheromone; Competence.
FT   PROPEP     1     24      POTENTIAL.
FT   CHAIN     25     41      COMPETENCE STIMULATING PEPTIDE.
SQ   SEQUENCE   41 AA;  4988 MW;  1276E5B6F72A1463 CRC64;

Query Match             24.4%;  Score 55.5;  DB 1;  Length 41;
Best Local Similarity   46.3%;  Pred. No. 0.82;
Matches   19;  Conservative   5;  Mismatches   14;  Indels   3;  Gaps   3;

Qy    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFPR-LF-NRS 39
          || :| ||| :||: |:| || | | ||:| || |: || ||
Db    1 MKNTEKLE-QFKEVTEAELQEIRGGDKRLPYFFKHLFSNRT 40


RESULT 6
VGLF_NDVB
ID   VGLF_NDVB     STANDARD;      PRT;   553 AA.
AC   P06156;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
DE   Fusion glycoprotein F1].
GN   F.
OS   Newcastle disease virus (strain Beaudette C/45) (NDV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11178;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87085486; PubMed=3025345;
RA   Chambers P., Millar N.S., Emmerson P.T.;
RT   "Nucleotide sequence of the gene encoding the fusion glycoprotein of
RT   Newcastle disease virus.";
RL   J. Gen. Virol. 67:2685-2694(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204898; PubMed=2705298;
RA   Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA   Nagai Y.;
RT   "Newcastle disease virus evolution. II. Lack of gene recombination in
RT   generating virulent and avirulent strains.";
RL   Virology 169:273-282(1989).
CC   -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC       MEMBRANES.
CC   -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC       LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC       FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute. There are no  restrictions on  its
CC   use by non-profit institutions as  long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04719; CAA28426.1; -.
DR   EMBL; M24697; AAA46648.1; -.
DR   EMBL; A03363; CAA00288.1; -.
DR   PIR; A27008; VGNZNV.
DR   PIR; F46329; F46329.
DR   HSSP; P04849; 1SVF.
DR   InterPro; IPR000776; Fusion_gly.
DR   Pfam; PF00523; fusion_gly; 1.
KW   Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW   Lipoprotein; Palmitate.
FT   SIGNAL      1     25
FT   CHAIN      26    553       FUSION GLYCOPROTEIN F0.
FT   CHAIN      26    116       F2 PROTEIN.
FT   CHAIN     117    553       F1 PROTEIN.
FT   TRANSMEM  117    136       POTENTIAL.
FT   DOMAIN    137    500       EXTRACELLULAR.
FT   TRANSMEM  501    527       POTENTIAL.
FT   DOMAIN    528    553       CYTOPLASMIC.
FT   CARBOHYD   85     85       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  366    366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  447    447       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID     523    523       PALMITATE (POTENTIAL).
SQ   SEQUENCE 553 AA; 59041 MW; 63BFD1692AFDE191 CRC64;

Query Match         24.4%;  Score 55.5;  DB 1;  Length 553;
Best Local Similarity  33.3%;  Pred. No. 15;
Matches  16;  Conservative  7;  Mismatches  20;  Indels  5;  Gaps  1;

Qy    4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
             |  |:||  : :|| |:   ||.||:  ||:| |||
Db  433 TLRLSGEFDATYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 7
VGLF_NDVH4
ID   VGLF_NDVH4    STANDARD;      PRT;   553 AA.
AC   P33613;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
DE   Fusion glycoprotein F1].
GN   F.
OS   Newcastle disease virus (strain B1-Hitchner/47) (NDV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11181;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204898; PubMed=2705298;
RA   Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA   Nagai Y.;
RT   "Newcastle disease virus evolution. II. Lack of gene recombination in
RT   generating virulent and avirulent strains.";
RL   Virology 169:273-282(1989).
CC   -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC       MEMBRANES.
CC   -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC   -!- LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC       FAMILY.
CC   ----------------------------------------------------------------------
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M24695; AAA46646.1; -.
DR   PIR; D46329; D46329.
DR   HSSP; P04849; 1SVF.
DR   InterPro; IPR000776; Fusion_gly.
DR   Pfam; PF00523; fusion_gly; 1.
KW   Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW   Lipoprotein; Palmitate.
FT   SIGNAL      1     25
FT   CHAIN      26    553       FUSION GLYCOPROTEIN F0.
FT   CHAIN      26    116       F2 PROTEIN.
FT   CHAIN     117    553       F1 PROTEIN.
FT   TRANSMEM  117    136       POTENTIAL.
FT   DOMAIN    137    500       EXTRACELLULAR.
FT   TRANSMEM  501    527       POTENTIAL.
FT   DOMAIN    528    553       CYTOPLASMIC.
FT   CARBOHYD   85     85       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  366    366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  447    447       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID     523    523       PALMITATE (POTENTIAL).
SQ   SEQUENCE 553 AA; 58909 MW; F163E9B7CFDE4493 CRC64;

Query Match         24.4%;  Score 55.5;  DB 1;  Length 553;
Best Local Similarity  33.3%;  Pred. No. 15;
Matches  16;  Conservative  7;  Mismatches  20;  Indels  5;  Gaps  1;

Qy    4 TLSLKNDF----KEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
             |  |:||  : :|| |:   ||.||:  ||:| |||
Db  433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 8
VGLF_NDVL
ID   VGLF_NDVL     STANDARD;      PRT;   553 AA.
AC   P33614;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
DE   Fusion glycoprotein F1].
GN   F.
OS   Newcastle disease virus (strain Las/46) (NDV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11184;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204898; PubMed=2705298;
RA   Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA   Nagai Y.;
RT   "Newcastle disease virus evolution. II. Lack of gene recombination in
RT   generating virulent and avirulent strains.";
RL   Virology 169:273-282(1989).
CC   -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC       MEMBRANES.
CC   -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC   -!- LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC       FAMILY.
CC   ----------------------------------------------------------------------
```

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M24696; AAA46647.1; -.
DR   PIR; E46329; E46329.
DR   HSSP; P04849; 1SVF.
DR   InterPro; IPR000776; Fusion_gly.
DR   Pfam; PF00523; fusion_gly; 1.
KW   Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW   Lipoprotein; Palmitate.
FT   SIGNAL         1     25
FT   CHAIN         26    553       FUSION GLYCOPROTEIN F0.
FT   CHAIN         26    116       F2 PROTEIN.
FT   CHAIN        117    553       F1 PROTEIN.
FT   TRANSMEM     117    136       POTENTIAL.
FT   DOMAIN       137    500       EXTRACELLULAR.
FT   TRANSMEM     501    527       POTENTIAL.
FT   DOMAIN       528    553       CYTOPLASMIC.
FT   CARBOHYD      85     85       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     366    366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     447    447       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID        523    523       PALMITATE (POTENTIAL).
SQ   SEQUENCE   553 AA;  59047 MW;  04E1lF38573B91DF CRC64;

Query Match             24.4%;  Score  55.5;  DB 1;  Length 553;
Best Local Similarity   33.3%;  Pred. No. 15;
Matches   16;  Conservative    7;  Mismatches   20;  Indels    5;  Gaps    1;

Qy    4 TLSLKNDF----KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
         |||  ||  :||:|| ||||| GGS ||   ||: :||  ||||:||
Db  433 TLRLSGEFDVTYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 9
VGLF_NDVTG
ID   VGLF_NDVTG     STANDARD;      PRT;   553 AA.
AC   P12571;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
DE   Fusion glycoprotein F1].
GN   F.
OS   Newcastle disease virus (strain Texas g.b./48) (NDV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11189;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88265873; PubMed=3388773;
RA   Schaper U.M., Fuller F.J., Ward M.D.W., Mehrotra Y., Stone H.O.,
RA   Stripp B.R., de Buysscher E.V.;
RT   "Nucleotide sequence of the envelope protein genes of a highly
RT   virulent, neurotropic strain of Newcastle disease virus.";
RL   Virology 165:291-295(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204898; PubMed=2705298;
RA   Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA   Nagai Y.;
RT   "Newcastle disease virus evolution. II. Lack of gene recombination in
RT   generating virulent and avirulent strains.";
RL   Virology 169:273-282(1989).
CC   -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC       MEMBRANES.
CC   -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC       LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PARAMXOVIRUSES FUSION GLYCOPROTEIN
CC       FAMILY.
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M23407; AAA46642.1; -.
DR   EMBL; M24698; AAA46649.1; -.
DR   PIR; B29201; VGNZGB.
DR   PIR; G46329; G46329.
DR   HSSP; P04849; 1SVF.
DR   InterPro; IPR000776; Fusion_gly.
DR   Pfam; PF00523; fusion_gly; 1.
KW   Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW   Lipoprotein; Palmitate.
FT   SIGNAL         1     25
FT   CHAIN         26    553       FUSION GLYCOPROTEIN F0.
FT   CHAIN         26    116       F2 PROTEIN.
FT   CHAIN        117    553       F1 PROTEIN.
FT   TRANSMEM     117    136       POTENTIAL.
FT   DOMAIN       137    500       EXTRACELLULAR.
FT   TRANSMEM     501    527       POTENTIAL.
FT   DOMAIN       528    553       CYTOPLASMIC.
FT   CARBOHYD      85     85       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     191    191       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     366    366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     447    447       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID        523    523       PALMITATE (POTENTIAL).
SQ   SEQUENCE   553 AA;  59017 MW;  5F537269FE76F9E9 CRC64;

Query Match             24.4%;  Score  55.5;  DB 1;  Length 553;
Best Local Similarity   33.3%;  Pred. No. 15;
Matches   16;  Conservative    7;  Mismatches   20;  Indels    5;  Gaps    1;

Qy    4 TLSLKNDF----KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
         |||  ||   :||:|| ||||| :||   ||: :||  ||||:|| |
Db  433 TLRLSGEFDATYQKNISIQDSQVIITGNLDISTELGNVNNSISNALNK 480

RESULT 10
CB32_YEAST
ID   CB32_YEAST     STANDARD;      PRT;   608 AA.
AC   P40969;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Centromere DNA-binding protein complex CBF3 subunit B (Centromere
DE   protein 3).
GN   CBF3B OR CEP3 OR YMR168C OR YM8520.17C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AC 502;
RX   MEDLINE=95045413; PubMed=7957085;
RA   Lechner J.;
RT   "A zinc finger protein, essential for chromosome segregation,
RT   constitutes a putative DNA binding subunit of the Saccharomyces
RT   cerevisiae kinetochore complex, Cbf3.";
RL   EMBO J. 13:5203-5211(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95181532; PubMed=7876302;
RA   Strunnikov A.V., Kingsbury J., Koshland D.;
RT   "CEP3 encodes a centromere protein of Saccharomyces cerevisiae.";
RL   J. Cell Biol. 128:749-760(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
```

RC   STRAIN=S288C / AB972;
RA   Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR CHROMOSOME SEGREGATION AND MOVEMENT OF
CC   CENTROMERES ALONG MICROTUBULES. IT PLAYS A ROLE IN THE ATTACHMENT
CC   OF CHROMOSOMES TO THE SPINDLE. CBF3 BINDS SELECTIVELY TO A HIGHLY
CC   CONSERVED DNA SEQUENCE CALLED CDEIII, FOUND IN CENTROMERS AND IN
CC   SEVERAL PROMOTERS.
CC   -!- SUBUNIT: CBF3 IS FORMED OF FOUR SUBUNITS, CBF3A (CBF2), CBF3B
CC   (CEP3), CBF3C (CTF13) AND CBF3D.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6) FUNGAL-TYPE BINUCLEAR
CC   CLUSTER DOMAIN.

DR   EMBL; X81396; CAA57159.1; -.
DR   EMBL; U12339; AAA57074.1; -.
DR   EMBL; Z49705; CAA89804.1; -.
DR   SGD; S0004778; CEP3.
DR   InterPro; IPR001138; Zn2_CY6_fungal.
DR   Pfam; PF00172; zn_clus; 1.
DR   SMART; SM00066; GAL4; 1.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW   Nuclear protein; Chromosomal protein; DNA-binding; Zinc;
KW   Metal-binding; Centromere.
FT   DNA_BIND   14   42   ZN(2)-CYS(6), FUNGAL-TYPE.
SQ   SEQUENCE   608 AA; 71358 MW; 2E24A050808OA09B CRC64;

Query Match   24.2%; Score 55; DB 1; Length 608;
Best Local Similarity   39.5%; Pred. No. 19;
Matches 15; Conservative   6; Mismatches 15; Indels   2; Gaps   1;

Qy   7 LKNDPK--EIKTDELEIIIGGSGSLSTFFRLFNRSFTQ 42
            |:|| :|:|:||: :|::|||| :|:|||:|
Db   541 LQNDIKIIELKNDEMFSLIKGLGSLVPLNKLRQESLLE 578

RESULT 11
HIS8_METJA
ID   HIS8_METJA   STANDARD;   PRT;   373 AA.
AC   Q58365;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole
DE   acetol-phosphate transaminase).
GN   HISC OR MJ0955.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";

RL   Science 273:1058-1073(1996).
CC   -!- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
CC   (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate (BY
CC   SIMILARITY).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC   AMINOTRANSFERASES.

DR   EMBL; U67539; AAB98960.1; -.
DR   TIGR; MJ0955; -.
DR   InterPro; IPR001511; Aminotran_1.
DR   InterPro; IPR001917; Aminotransf_2.
DR   Pfam; PF00155; aminotran_1_2; 1.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; FALSE_NEG.
KW   Histidine biosynthesis; Transferase; Aminotransferase;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING   231   231   PYRIDOXAL PHOSPHATE (PROBABLE).
SQ   SEQUENCE   373 AA; 42961 MW; 369DB5EB4C671218 CRC64;

Query Match   24.0%; Score 54.5; DB 1; Length 373;
Best Local Similarity   39.4%; Pred. No. 13;
Matches 13; Conservative   5; Mismatches 12; Indels   3; Gaps   1;

Qy   7 LKNDFKEIKTDELEIIIGGSGSS---LSTFFRLF 36
          :| :|| :|| |||| :|:||| : :|||
Db   79 MKELSKFLNVDEENIIVGGDGADEIIDIFIRTF 111

RESULT 12
DNAA_RICCN
ID   DNAA_RICCN   STANDARD;   PRT;   463 AA.
AC   Q92H56;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Chromosomal replication initiator protein dnaA.
GN   DNAA OR RC0916.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
CC   -!- FUNCTION: Plays an important role in the initiation and regulation
CC   of chromosomal replication. Binds to the origin of replication; it
CC   binds specifically double-stranded DNA at a 9 bp consensus (dnaA
CC   box): 5'-TTATC(C/A)A(C/A)A-3'. DnaA binds to ATP and to acidic
CC   phospholipids (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE DNAA FAMILY.

```
CC   ---------
DR   EMBL; AE008645; AAL03454.1; -.
DR   InterPro; IPR001957; Bac_DnaA.
DR   Pfam; PF00308; bac_dnaA; 1.
DR   PROSITE; PS01008; DNAA; 1.
KW   DNA replication; DNA-binding; ATP-binding; Complete proteome.
FT   NP_BIND   168    175      ATP (POTENTIAL).
SQ   SEQUENCE   463 AA;  52942 MW;  B48D2FE43A70A8EA CRC64;

Query Match       24.08%;  Score 54.5;  DB 1;  Length 463;
Best Local Similarity   28.2%;  Pred. No. 16;
Matches  11;  Conservative  8;  Mismatches  17;  Indels  3;  Gaps  1;

Qy    2 KKTLSLKNDFKEIKT---DELEIIIGSGSLSTFFRLFN 37
          | :|: | :|:   |||||| |||| |
Db  213 KEVISFKFKFKFRSVDVLMIDDIQFICGKDSTQEEFFHTFN 251
```

RESULT 13
SYFB_PORPU

```
ID   SYFB_PORPU     STANDARD;     PRT;   720 AA.
AC   P51346;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE   tRNA ligase beta chain) (PheRS).
GN   PHET
OS   Porphyra purpurea.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX   NCBI_TaxID=2787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RA   Reith M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT   genome.";
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC       diphosphate + L-phenylalanyl-tRNA(Phe).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC       CHAIN FAMILY. SUBFAMILY 1.
CC   ---------
```

```
DR   EMBL; U38804; AAC08232.1; -.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Chloroplast.
SQ   SEQUENCE   720 AA;  82592 MW;  495E49947117F62A CRC64;

Query Match       24.08%;  Score 54.5;  DB 1;  Length 720;
Best Local Similarity   31.58%;  Pred. No. 26;
Matches  17;  Conservative  7;  Mismatches  9;  Indels  21;  Gaps  2;

Qy    6 SLKNDFKEIKTDELEIIIGGS------------------GSLSTFFRLFNRS 39
          :|| : | :|: :| :|:!||                     ||| ||:|
Db  495 NLKRN-KIIETKLAIILGGSLDIRSEWSEPAHSLNWYEAKGIIENFFRKLNKS 547
```

RESULT 14
CXB2_MOUSE

```
ID   CXB2_MOUSE     STANDARD;     PRT;   226 AA.
AC   Q00977;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Gap junction beta-2 protein (Connexin 26) (Cx26).
GN   GJB2 OR CXN-26.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92354576; PubMed=1322820;
RA   Willecke K., Nicholson B.J., Dahl E., Kozjek G., Hennemann H.;
RT   "Molecular cloning of mouse connexins26 and -32: similar genomic
RT   organization but distinct promoter sequences of two gap junction
RT   genes.";
RL   Eur. J. Cell Biol. 58:81-89(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91285228; PubMed=2060697;
RA   Nishi M., Kumar N.M., Gilula N.B.;
RT   "Developmental regulation of gap junction gene expression during
RT   mouse embryonic development.";
RL   Dev. Biol. 146:117-130(1991).
CC   -!- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
CC       PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
CC       MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
CC   -!- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: LIVER, KIDNEY, INTESTINE, LUNG, SPLEEN,
CC       STOMACH, TESTIS, & BRAIN, BUT NOT HEART & ADULT SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. BETA-TYPE (GROUP I)
CC       SUBFAMILY.
CC   ---------
```

```
DR   EMBL; M81445; AAA37495.1; -.
DR   EMBL; M63803; AAA37276.1; -.
DR   PIR; C49769; C49769.
DR   MGD; MGI:95720; Gjb2.
DR   InterPro; IPR000500; Connexin.
DR   Pfam; PF00029; connexin; 1.
DR   PRINTS; PR00206; CONNEXIN.
DR   SMART; SM00037; CNX; 1.
DR   PROSITE; PS00407; CONNEXINS_1; 1.
DR   PROSITE; PS00408; CONNEXINS_2; 1.
KW   Gap junction; Transmembrane.
FT   DOMAIN       1     19      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    20     40      POTENTIAL.
FT   DOMAIN      41     75      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    76     96      POTENTIAL.
FT   DOMAIN      97    143      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   144    164      POTENTIAL.
FT   DOMAIN     165    189      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   190    210      POTENTIAL.
FT   DOMAIN     211    226      CYTOPLASMIC (POTENTIAL).
FT   CONFLICT    68     68      H -> Y (IN REF. 2).
FT   CONFLICT   215    215      V -> I (IN REF. 2).
SQ   SEQUENCE   226 AA;  26411 MW;  0D8178F7A39E0B6 CRC64;

Query Match       23.8%;  Score 54;  DB 1;  Length 226;
Best Local Similarity   33.9%;  Pred. No. 8.5;
Matches  21;  Conservative  6;  Mismatches  9;  Indels  26;  Gaps  4;

Qy    7 LKNDFK--EIKTDELEIIIGGSGSL-----STFFR---------LFNRSFTQAL 44
          |||:||  :|::| : | |||| |   |||         ::| |||
Db  111 IKNEFKDIEEIKTQKVRI----EGSLWTYTTSIFFRVIFEAVFMVFYIMYNGFFMQRL 166
```

Search completed: November   5, 2002, 10:57:52
Job time : 13.9851 secs

```
Qy     45 GK 46
           |
Db    167 VK 168


RESULT 15
CXB2_RAT
ID   CXB2_RAT        STANDARD;        PRT;     226 AA.
AC   P21994;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Gap junction beta-2 protein (Connexin 26) (Cx26).
GN   GJB2 OR CXN-26.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90094541; PubMed=2557354;
RA   Zhang J.T., Nicholson B.J.;
RT   "Sequence and tissue distribution of a second protein of hepatic gap
RT   junctions, Cx26, as deduced from its cDNA.";
RL   J. Cell Biol. 109:3391-3401(1989).
RN   [2]
RP   SEQUENCE OF 1-18.
RX   MEDLINE=88039049; PubMed=2823143;
RA   Nicholson B.J., Dermietzel R., Teplow D., Traub O., Willecke K.,
RA   Revel J.-P.;
RT   "Two homologous protein components of hepatic gap junctions.";
RL   Nature 329:732-734(1987).
CC   -!- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
CC       PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
CC       MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
CC   -!- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: LIVER, KIDNEY, INTESTINE, LUNG, SPLEEN,
CC       STOMACH, TESTIS, & BRAIN, BUT NOT HEART & ADULT SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. BETA-TYPE (GROUP I)
CC       SUBFAMILY.
DR   PIR; A33646; A33646.
DR   InterPro; IPR000500; Connexin.
DR   Pfam; PF00029; connexin; 1.
DR   PRINTS; PR00206; CONNEXIN.
DR   SMART; SM00037; CNX; 1.
DR   PROSITE; PS00407; CONNEXINS_1; 1.
DR   PROSITE; PS00408; CONNEXINS_2; 1.
KW   Gap junction; Transmembrane.
FT   DOMAIN        1     19       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     20     40       POTENTIAL.
FT   DOMAIN       41     75       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     76     96       POTENTIAL.
FT   DOMAIN       97    143       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    144    164       POTENTIAL.
FT   DOMAIN      165    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    190    210       POTENTIAL.
FT   DOMAIN      211    226       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   226 AA;  26451 MW;  E36EC95F1B235A0D CRC64;


Query Match             23.8%;  Score 54;  DB 1;  Length 226;
Best Local Similarity   33.9%;  Pred. No. 8.5;
Matches    21;  Conservative    6;  Mismatches    9;  Indels   26;  Gaps     4;

Qy      7 LKNDFK--EIKTDELEIIIGGSGSL-----STFFR----------LFNRSFTQAL 44
          ::||:||  ||||.:||: :|  |     |  |||          ::| :|||
Db    111 IKNEFKDIEEIKTQKVRI----EGSLWWTYTTSIFFRVIFEAVFMYVFIMYNGFFMQRL 166


Qy     45 GK 46
          |
Db    167 VK 168
```

```
          GenCore version 5.1.3
          Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      November  5, 2002, 10:55:02 ; Search time 29.5224 Seconds
                                  (without alignments)
                        269.550 Million cell updates/sec

Title:       US-09-833-017B-2
Perfect score: 227
Sequence:    1 MKKTLSLKNDFKFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:      562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_19:*
             1:  sp_archea:*
             2:  sp_bacteria:*
             3:  sp_fungi:*
             4:  sp_human:*
             5:  sp_invertebrate:*
             6:  sp_mammal:*
             7:  sp_mhc:*
             8:  sp_organelle:*
             9:  sp_phage:*
             10: sp_plant:*
             11: sp_rodent:*
             12: sp_virus:*
             13: sp_vertebrate:*
             14: sp_unclassified:*
             15: sp_rvirus:*
             16: sp_bacteriap:*
             17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	227	100.0	46	2	Q99QI5	streptococc
2	220	96.9	46	2	Q9APK7	streptococc
3	209	92.1	43	2	Q9APK6	streptococc
4	69	30.4	850	16	Q97N40	streptococc
5	63.5	28.0	648	2	Q9L651	lactococcus
6	59.5	26.2	275	16	Q9CPE8	pasteurella
7	58	25.6	48	16	O33596	streptococc
8	58	25.6	1009	16	Q99WU3	staphylococc
9	57.5	25.3	453	3	Q9C1A3	saccharomyc
10	57.5	25.3	516	10	Q9SUK0	arabidopsis
11	57.5	25.3	516	10	Q9LK23	arabidopsis
12	57	25.1	591	2	Q9AHY9	photorhabdu
13	56.5	24.9	407	2	Q9RQ06	lactococcus
14	56.5	24.9	407	2	Q91931	lactococcus
15	56.5	24.9	408	16	Q9KIF7	lactococcus
16	56.5	24.9	593	16	Q92GD6	rickettsia
17	56	24.7	1250	16	Q9PPL7	campylobact
18	55.5	24.4	125	12	O41368	newcastle d
19	55.5	24.4	125	12	O41472	newcastle d
20	55.5	24.4	416	16	Q92EB0	listeria in
21	55.5	24.4	553	12	O90339	newcastle d
22	55.5	24.4	553	12	Q9DLD4	newcastle d
23	55.5	24.4	553	12	Q9WMH7	newcastle d
24	55.5	24.4	553	12	O92150	newcastle d
25	55.5	24.4	553	12	Q9WLE2	newcastle d
26	55.5	24.4	553	12	Q9WLE1	newcastle d
27	55.5	24.4	553	12	Q91HX4	newcastle d
28	55.5	24.4	553	12	Q91AH8	newcastle d
29	55	24.4	837	10	Q9M819	arabidopsis
30	55	24.4	1304	2	Q9U0H4	plasmodium
31	55	24.2	48	2	O33615	streptococc
32	54.5	24.0	230	16	Q9CGX0	lactococcus
33	54.5	24.0	255	2	Q9AKI3	rickettsia
34	54.5	24.0	255	2	Q9AKN6	rickettsia
35	54.5	24.0	398	16	Q9A1N0	streptococc
36	54.5	24.0	463	16	Q92H56	rickettsia
37	54.5	24.0	819	5	O97330	plasmodium
38	54	23.8	48	2	O33651	streptococc
39	54	23.8	235	2	Q9X9J9	vibrio para
40	54	23.8	243	16	Q97M22	clostridium
41	54	23.8	2353	2	P71401	haemophilus
42	53.5	23.6	125	12	O41386	newcastle d
43	53.5	23.6	125	12	O41388	newcastle d
44	53.5	23.6	165	1	Q00392	methanococc
45	53.5	23.6	176	2	Q9RJ74	streptomyce

ALIGNMENTS

```
RESULT 1
Q99QI5
ID   Q99QI5      PRELIMINARY;      PRT;       46 AA.
AC   Q99QI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN PRECURSOR.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB14, H7, LT11, NG8, AND UA159;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277152; AAK01542.1; -.
DR   EMBL; AF277153; AAK01543.1; -.
DR   EMBL; AF277155; AAK01545.1; -.
DR   EMBL; AF277156; AAK01546.1; -.
DR   EMBL; AF277157; AAK01547.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
FT   CHAIN      26     46       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;   5211 MW;  38FA62B6F78FC3BF CRC64;

Query Match          100.0%;    Score 227;   DB 2;   Length 46;
Best Local Similarity 100.0%;    Pred. No. 5.3e-22;
Matches   46;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MKKTLSLKNDFKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MKKTLSLKNDFKFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
```

```
RESULT 2
Q9APK7
ID   Q9APK7      PRELIMINARY;      PRT;    46 AA.
AC   Q9APK7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BM71;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277151; AAK01541.1; -.
FT   CHAIN        26     46       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;  5195 MW;  38E0B9DB5B8FC3BF CRC64;

Query Match             96.9%;  Score 220;  DB 2;  Length 46;
Best Local Similarity   97.8%;
  Matches   45;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
              ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKKTPSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46

RESULT 3
Q9APK6
ID   Q9APK6      PRELIMINARY;      PRT;    43 AA.
AC   Q9APK6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH1005;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277154; AAK01544.1; -.
FT   CHAIN        26     43       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   43 AA;  4927 MW;  E6A78FC3BF6156C7 CRC64;

Query Match             92.1%;  Score 209;  DB 2;  Length 43;
Best Local Similarity   97.7%;  Pred. No. 1e-19;
  Matches   42;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQA 43
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 MKKTLSLKNDFKEIKTDELEIIIGGSGTLSTFFRLFNRSFTQA 43

RESULT 4
Q97N40
ID   Q97N40      PRELIMINARY;      PRT;    850 AA.
AC   Q97N40;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
GN   SP2231.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae";
RL   Science 293:498-506(2001).
DR   EMBL; AE007510; AAK76279.1; -.
DR   TIGR; SP2231; -.
KW   Complete proteome.
SQ   SEQUENCE   850 AA;  97303 MW;  1ADED613F06B5115 CRC64;

Query Match             30.4%;  Score 69;  DB 16;  Length 850;
Best Local Similarity   33.3%;  Pred. No. 2.5;
  Matches   19;  Conservative   10;  Mismatches   16;  Indels   12;  Gaps    2;

Qy          2 KKTLSLKNDF------KEIKTDELEIIIGGSGSLS---TFFRLFNRSFTQALGK 46
              |||  :|| ||| :  :| ::|::|:|        |    :| :: :|:|: |
Db        692 KKTFTTDNVFSFFNLGYTKEKKTPNINVSFPGNSQVSFESPTFYRLDTKTFTEAIQK 748

RESULT 5
Q9L651
ID   Q9L651      PRELIMINARY;      PRT;    648 AA.
AC   Q9L651;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL LACTOCOCCIN 972 IMMUNITY PROTEIN (LACTIS PLASMID PBL1 DNA
DE   FOR LACTOCOCCIN 972 OPERON).
GN   LCLB.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG   Plasmid pBL1.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IPLA 972; PLASMID=PBL1;
RA   Sanchez C., Hernandez de Rojas A., Martinez B., Arguelles M.E.,
RA   Suarez J.E., Rodriguez A., Mayo B.;
RT   "Nucleotide sequence of pBL1, a bacteriocin-producing plasmid from
RT   Lactococcus lactis IPLA 972.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IPLA 972; PLASMID=PBL1;
RA   Suarez J.E.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IPLA 972; PLASMID=PBL1;
RX   MEDLINE=20055640; PubMed=10589723;
RA   Martinez B., Fernandez M., Rodriguez A., Suarez J.E.;
RT   "Synthesis of lactococcin 972, a bacteriocin produced by Lactococcus
RT   lactis IPLA 972, depends on the expression of a plasmid-encoded
```

RT   bicistronic operon.";
RL   Microbiology 145:3155-3161(1999).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IPLA 972; PLASMID=PBL1;
RA   Martinez B.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF242367; AAF64055.1; -.
DR   EMBL; AJ002203; CAC03468.1; -.
KW   Plasmid.
SQ   SEQUENCE   648 AA;   74139 MW; 79530E501D6371F7 CRC64;

Query Match   28.0%; Score 63.5; DB 2; Length 648;
Best Local Similarity   46.2%; Pred. No. 9.6;
Matches 18; Conservative   8; Mismatches 10; Indels 3; Gaps 2;

Qy   5 LSLKNDFKEIKTDELEIIIGGSGSL-STF-FRLFNRSF 40
          |:|| ||| ||| :| ||| | | :| || |
Db 528 LSVKNRFVEIKSQQLFIVVTSSIALISSTFLFYLMNKIY 566

RESULT 6
Q9CPE8
ID   Q9CPE8   PRELIMINARY;   PRT; 275 AA.
AC   Q9CPE8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROC.
GN   PROC OR PM0095.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006044; AAK02179.1; -.
DR   InterPro; IPR000304; P5CR.
DR   Pfam; PF01089; P5CR; 1.
DR   PROSITE; PS00521; P5CR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   275 AA;   29556 MW; 01IC7F1AD31A7D7 CRC64;

Query Match   26.2%; Score 59.5; DB 16; Length 275;
Best Local Similarity   29.6%; Pred. No. 12;
Matches 16; Conservative   7; Mismatches 16; Indels 15; Gaps 1;

Qy   6 SLKNDFKEIKTD---------ELEIIIGGSGSLSTFFRLFNRSFTQAL 44
          ||| :|: | : |||| | |||| :| ||
Db 141 SLKPEYKQFTQDLLNAVGKTCWVTQEADMHTIFAGSGSSPAYFLFMEAMQQAL 194

RESULT 7
O33596
ID   O33596   PRELIMINARY;   PRT; 48 AA.
AC   O33596;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PEPTIDE PRECURSOR.
GN   COMC.
OS   Streptococcus anginosus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1328;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10713;
RX   MEDLINE=9801953; PubMed=9352904;
RA   Havarstein L.S., Hakenbeck R., Gaustad P.;
RT   "Natural competence in the genus Streptococcus: Evidence that
RT   streptococci can change phenotype by interspecies recombinational
RT   exchanges.";
RL   J. Bacteriol. 179:6589-6594(1997).
DR   EMBL; AJ000864; CAA04341.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
KW   Signal.
FT   SIGNAL   1   32
FT   CHAIN   33   48   COMPETENCE STIMULATING PEPTIDE.
SQ   SEQUENCE   48 AA;   5621 MW; EB6DF7AEDE4E94DB CRC64;

Query Match   25.6%; Score 58; DB 2; Length 48;
Best Local Similarity   30.2%; Pred. No. 2.8;
Matches 16; Conservative   9; Mismatches 16; Indels 12; Gaps 2;

Qy   1 MKKTLSLKN------DFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
          ||| ||||| |||| ::|:|||| : :|||| ||
Db   1 MKKLFAKKEVKKEVVEFKELNDEQLDKIIGGDSRIRMGF------DFSKLFGK 48

RESULT 8
Q99WU3
ID   Q99WU3   PRELIMINARY;   PRT; 1009 AA.
AC   Q99WU3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SA0272 PROTEIN (HYPOTHETICAL PROTEIN SAV0283).
GN   SA0272 OR SAV0283.
OS   Staphylococcus aureus (strain N315), and
OS   Staphylococcus aureus (strain Mu50).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group: Staphylococcus.
OX   NCBI_TaxID=158879, 158878;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami A., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima A., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357;1225-1240(2001).
DR   EMBL; AP003130; BABA1496.1; -.
DR   EMBL; AP003358; BAB56445.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   1009 AA; 114781 MW; 16AD06D170CA31FE CRC64;

Query Match   25.6%; Score 58; DB 16; Length 1009;
Best Local Similarity   34.3%; Pred. No. 79;
Matches 12; Conservative   8; Mismatches 15; Indels 0; Gaps 0;

Qy   9 NDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQA 43
          || |:|| ||| : : :|: :|: :|:
Db 202 NDFPELFTDTLVNSISANKDITKWFQTYNKSLLSA 236

RESULT 9
Q9C1A3
ID   Q9C1A3   PRELIMINARY;   PRT; 453 AA.
AC   Q9C1A3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   SULFATE TRANSPORTER SUL2-LA (FRAGMENT).

```
OS   Saccharomyces pastorianus (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=27292;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   James A.B., Slaughter C., Meaden P.G.;
RT   "Characterization of partial coding sulfate transporter sequences from
RT   Saccharomyces pastorianus and Saccharomyces bayanus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF364410; AAK32879.1; -.
DR   InterPro; IPR001902; Sulfate_transp.
DR   Pfam; PF00916; Sulfate_transp; 1.
FT   NON_TER       1      1
FT   NON_TER     453    453
SQ   SEQUENCE   453 AA;  49519 MW;  137117A0D12BD31F CRC64;

Query Match             25.3%;  Score 57.5;  DB 3;  Length 453;
Best Local Similarity   36.7%;          Pred. No. 38;
Matches  18;  Conservative   6;  Mismatches  18;  Indels   7;  Gaps   2;

Qy     1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRL------FNRSFTQA 43
           ::: :   ::::  || ||||| ||||  |:|| :|| :|
Db   227 ISKSFGRVNDYKVVPDQEL-IAIGVSNLLGTFFNAYPATGSFSRSALKA 274


RESULT 10
Q9SUK0
ID   Q9SUK0        PRELIMINARY;       PRT;   516 AA.
AC   Q9SUK0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
GN   ACG9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99364543; PubMed=10437832;
RA   Wendt U.K., Hauschild R., Lange C., Pietersma M., Wenderoth I.,
RA   von Schaewen A.;
RT   "Evidence for functional convergence of redox regulation in G6PDH
RT   isoforms of cyanobacteria and higher plants.";
RL   Plant Mol. Biol. 40:487-494(1999).
CC   -!- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONO-
CC       DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC   -!- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AJ010970; CAB52674.1; -.
DR   HSSP; P11411; 1DPG.
DR   InterPro; IPR001282; G6PD.
DR   Pfam; PF00479; G6PD; 1.
DR   Pfam; PF02781; G6PD_C; 1.
DR   PRINTS; PR00079; G6PDHDRGNASE.
DR   ProDom; PD001129; G6PD; 1.
DR   PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW   Glucose metabolism; NADP; Oxidoreductase.
SQ   SEQUENCE   516 AA;  59171 MW;  A7623780F053CAIC CRC64;

Query Match             25.3%;  Score 57.5;  DB 10;  Length 516;
Best Local Similarity   41.5%;          Pred. No. 44;
Matches  22;  Conservative   7;  Mismatches  11;  Indels  13;  Gaps   5;

Qy     1 MKKTLSLKND-FKE----IKTDELEIII-GGSGGSLS---TF---FRLFNRSF 40
           |:| :|||| :||     :|||||||  | ::||:| :  :|   |:| |||
Db     8 MEKRSTLKNDSFVKEYNPITETGSLSIIVLGASGDLAKKKTFPALFNLFHQGF 60
```

```
RESULT 11
Q9LK23
ID   Q9LK23        PRELIMINARY;       PRT;   516 AA.
AC   Q9LK23;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
CC   -!- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONO-
CC       DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC   -!- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AP000381; BAB02125.1; -.
DR   HSSP; P11411; 1DPG.
DR   InterPro; IPR001282; G6PD.
DR   Pfam; PF00479; G6PD; 1.
DR   Pfam; PF02781; G6PD_C; 1.
DR   PRINTS; PR00079; G6PDHDRGNASE.
DR   ProDom; PD001129; G6PD; 1.
DR   PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW   Glucose metabolism; NADP; Oxidoreductase.
SQ   SEQUENCE   516 AA;  59157 MW;  407E42590EFC20E2 CRC64;

Query Match             25.3%;  Score 57.5;  DB 10;  Length 516;
Best Local Similarity   41.5%;          Pred. No. 44;
Matches  22;  Conservative   7;  Mismatches  11;  Indels  13;  Gaps   5;

Qy     1 MKKTLSLKND-FKE----IKTDELEIII-GGSGGSLS---TF---FRLFNRSF 40
           |:| :|||| :||     :|||||||  | ::||:| :  :|   |:| |||
Db     8 MEKRSTLKNDSFVKEYNPVTETGSLSIIVLGASGDLAKKKTFPALFNLFHQGF 60


RESULT 12
Q9AHY9
ID   Q9AHY9        PRELIMINARY;       PRT;   591 AA.
AC   Q9AHY9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   YFAA (FRAGMENT).
GN   YFAA.
OS   Photorhabdus luminescens (Xenorhabdus luminescens).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Photorhabdus.
OX   NCBI_TaxID=29488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NC19, ATCC29304;
RX   MEDLINE=21225535; PubMed=11325940;
RA   Ciche T.A., Bintrim S.B., Horswill A.R., Ensign J.C.;
RT   "A Phosphopantetheinyl Transferase Homolog Is Essential for
RT   Photorhabdus luminescens To Support Growth and Reproduction of the
RT   Entomopathogenic Nematode Heterorhabditis bacteriophora.";
```

```
RL   J. Bacteriol. 183:3117-3126(2001).
DR   EMBL; AF288085; AAK16098.1; -.
DR   HSSP; P39435; 1KAS.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
FT   NON_TER      591    591
SQ   SEQUENCE   591 AA;  63878 MW;  AF1451C6D500891 CRC64;

Query Match             25.1%;  Score  57;  DB  2;  Length 591;
Best Local Similarity   34.3%;  Pred. No. 59;
Matches   12;  Conservative    7;  Mismatches   16;  Indels    0;  Gaps     0;

Qy         12 KEIKTDELEIIGGSGSLSTFFRLFNRSFTQALGK 46
              :| |::|::||:|::|||:|:||:||
Db        149 QRVNRDRIGVILGGGNGNTLFSLAGRQQTPYLRK 183

RESULT 13
Q9RQ06
ID   Q9RQ06      PRELIMINARY;      PRT;   407 AA.
AC   Q9RQ06;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BUSAA.
GN   BUSAA.
OS   Lactococcus lactis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1358;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCDO 763;
RX   MEDLINE=99445450; PubMed=10515910;
RA   Obis D., Guillot A., Gripon J.C., Renault P., Bolotin A., Mistou M.Y.;
RT   "Genetic and biochemical characterization of a high-affinity betaine
RT   uptake system (BusA) in Lactococcus lactis reveals a new functional
RT   organization within bacterial ABC transporters.";
RL   J. Bacteriol. 181:6238-6246(1999).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AF139575; AAF04258.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR000644; CBS.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00116; CBS; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport.
SQ   SEQUENCE   407 AA;  45696 MW;  D5DFBC0F8764B3A9 CRC64;

Query Match             24.9%;  Score  56.5;  DB  2;  Length 407;
Best Local Similarity   45.7%;  Pred. No. 46;
Matches   16;  Conservative    3;  Mismatches   15;  Indels    1;  Gaps     1;

Qy         13 EIKTDELEIIGGSGGS-LSTFFRLFNRSFTQALGK 46
              || |::||:|| ||| || || |||||||||
Db         50 EINEGEIFVIMGLSGSGKSTLLRLLNRLIEPTSGK 84

RESULT 14
Q9L931
ID   Q9L931      PRELIMINARY;      PRT;   407 AA.
AC   Q9L931;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   OPUAA.
GN   OPUAA.
OS   Lactococcus lactis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1358;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MG1363;
RX   MEDLINE=20393175; PubMed=10939245;
RA   Bouvier J., Bordes P., Romeo Y., Fourcans A., Bouvier I.,
RA   Gutierrez C.;
RT   "Characterization of OpuA, a glycine-betaine uptake system of
RT   Lactococcus lactis.";
RL   J. Mol. Microbiol. Biotechnol. 2:199-205(2000).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AF184955; AAF40473.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR000644; CBS.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00116; CBS; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport.
SQ   SEQUENCE   407 AA;  45653 MW;  656D9097AE58407F CRC64;

Query Match             24.9%;  Score  56.5;  DB  2;  Length 407;
Best Local Similarity   45.7%;  Pred. No. 46;
Matches   16;  Conservative    3;  Mismatches   15;  Indels    1;  Gaps     1;

Qy         13 EIKTDELEIIGGSGGS-LSTFFRLFNRSFTQALGK 46
              || |::||:|| ||| || || |||||||||
Db         50 EINEGEIFVIMGLSGSGKSTLLRLLNRLIEPTSGK 84

RESULT 15
Q9KIF7
ID   Q9KIF7      PRELIMINARY;      PRT;   408 AA.
AC   Q9KIF7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   OPUAA (BETAINE ABC TRANSPORTER ATP BINDING PROTEIN).
GN   OPUAA OR BUSAA.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20318987; PubMed=10860977;
RA   van der Heide T., Poolman B.;
RT   "Osmoregulated ABC-transport system of Lactococcus lactis senses water
RT   stress via changes in the physical state of the membrane.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:7102-7106(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS).
DR   EMBL; AF234619; AAF37878.1; -.
DR   EMBL; AE006375; AAK05550.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR000644; CBS.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF00571; CBS; 2.
```

```
DR    SMART; SM00382; AAA; 1.
DR    SMART; SM00116; CBS; 2.
DR    PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW    ATP-binding; Complete proteome; Transport.
SQ    SEQUENCE  408 AA;  45750 MW;  EBBA8A87BE563B88 CRC64;

Query Match           24.9%;  Score 56.5;  DB 16;  Length 408;
Best Local Similarity 45.7%;  Pred. No. 46;
Matches  16;  Conservative  3;  Mismatches  15;  Indels  1;  Gaps  1;

Qy    13 EIKTDELEIIIGGSGS-LSTFFRLFNRSFTQALGK 46
          ||    |: :|:| ||| :|  || ||        ||
Db    50 EINEGEIFVIMGLSGSGKSTLLRLLNRLLIEPTSGK 84

Search completed: November  5, 2002, 10:57:30
Job time : 34.5224 secs
```

OM protein - protein search, using sw model

```
                GenCore version 5.1.3
      Copyright (c) 1993 - 2002 Compugen Ltd.

Run on:    November 5, 2002, 10:57:42 ; Search time 36.3881 Seconds
                                        (without alignments)
                           140.414 Million cell updates/sec

Title:           US-09-833-017B-2
Perfect score:   46
Sequence:      1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        747574 seqs, 111073796 residues
Word size :      0
Total number of hits satisfying chosen parameters:          747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 1000 summaries
Database :  A_Geneseq_032802:*
    1:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT:*
    2:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT:*
    3:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT:*
    4:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT:*
    5:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT:*
    6:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT:*
    7:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT:*
    8:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT:*
    9:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT:*
   10:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT:*
   11:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT:*
   12:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT:*
   13:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT:*
   14:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT:*
   15:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT:*
   16:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT:*
   17:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT:*
   18:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT:*
   19:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT:*
   20:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:*
   21:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT:*
   22:   /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	15.2	115	22	AA00376	Human polypeptide
2	7	15.2	118	22	AA01734	Human polypeptide
3	7	15.2	126	22	AA00422	Human polypeptide
4	7	15.2	150	22	AA01182	Human polypeptide
5	7	15.2	186	21	AAG19904	Arabidopsis thalia
6	7	15.2	257	22	AAG19903	Arabidopsis thalia
7	7	15.2	287	21	AAG19902	Arabidopsis thalia
8	7	15.2	558	14	AAR42086	Mouse RelB protein
9	7	15.2	586	21	AAG30638	Arabidopsis thalia
10	7	15.2	612	21	AAG30637	Arabidopsis thalia
11	7	15.2	618	21	AAG30636	Arabidopsis thalia
12	7	15.2	658	22	ABG17981	Novel human diagno
13	7	15.2	767	22	ABB58240	Drosophila melanog
14	7	15.2	929	22	ABB70857	Drosophila melanog
15	7	15.2	2771	22	ABB63880	Drosophila melanog
16	6	13.0	17	16	AAR67567	Antagonist (18) of
17	6	13.0	19	21	AAY85614	Cathepsin-D antige
18	6	13.0	20	18	AAW33042	Human heat shock p
19	6	13.0	20	18	AAW12350	Human hsp60 peptid
20	6	13.0	27	20	AAY40080	Peptide sequence d
21	6	13.0	30	22	ABB29081	Peptide #1732 enco
22	6	13.0	30	22	ABB34240	Peptide #1746 enco
23	6	13.0	30	22	ABB19676	Protein #1675 enco
24	6	13.0	30	22	AAM55035	Human brain expres
25	6	13.0	30	22	AAM67423	Human bone marrow
26	6	13.0	30	22	AAM15251	Peptide #1685 enco
27	6	13.0	30	22	AAM27712	Peptide #1749 enco
28	6	13.0	30	22	AAM02995	Peptide #1677 enco
29	6	13.0	34	22	AAM91358	Human immune/haema
30	6	13.0	48	22	AAM23675	Human EST encoded
31	6	13.0	51	17	AAR94978	Calcium channel is
32	6	13.0	51	17	AAR94979	Calcium channel is
33	6	13.0	54	22	AAR77003	Human colon cancer
34	6	13.0	61	22	AAU46669	Propionibacterium
35	6	13.0	62	17	AAR97851	Rat brain calcium
36	6	13.0	66	17	AAR97852	Rat brain calcium
37	6	13.0	66	20	AAY60064	Human endometrium
38	6	13.0	66	22	ABG04703	Novel human diagno
39	6	13.0	69	22	ABB32283	Peptide #4934 enco
40	6	13.0	69	22	ABB22836	Protein #4835 enco
41	6	13.0	69	22	AAM58198	Human brain expres
42	6	13.0	69	22	AAM70656	Human bone marrow
43	6	13.0	69	22	AAM30969	Peptide #5006 enco
44	6	13.0	70	22	AAM92947	Human digestive sy
45	6	13.0	70	22	AAU20043	Human liver associ
46	6	13.0	72	22	AAO13104	Human polypeptide
47	6	13.0	78	22	AAM89377	Human immune/haema
48	6	13.0	83	22	ABB16579	Human nervous syst
49	6	13.0	96	19	AAW59796	Nucleotide sequenc
50	6	13.0	96	22	AAU56723	Propionibacterium
51	6	13.0	97	22	AAU45726	Propionibacterium
52	6	13.0	97	22	AAU46334	Propionibacterium
53	6	13.0	101	21	AAB57131	Human prostate can
54	6	13.0	105	22	ABG13791	Novel human diagno
55	6	13.0	109	22	AAO09578	Human polypeptide
56	6	13.0	110	20	AAY12436	Human 5' EST secre
57	6	13.0	111	22	AAO10094	Human polypeptide
58	6	13.0	112	22	AAU43924	Propionibacterium
59	6	13.0	118	22	AAO01659	Human polypeptide
60	6	13.0	118	22	AAU02075	Synthetic autoanti
61	6	13.0	121	22	AAO06498	Human polypeptide
62	6	13.0	122	20	AAY400074	Peptide sequence d
63	6	13.0	126	21	AAG08937	Arabidopsis thalia
64	6	13.0	132	18	AAU20491	H. pylori cytoplas
65	6	13.0	132	22	AAU20354	Human secreted pro
66	6	13.0	139	22	AAB62816	E. canis lipoprote
67	6	13.0	141	20	AAY35818	Amino acid sequenc
68	6	13.0	141	20	AAU45360	Propionibacterium
69	6	13.0	147	22	AAE17946	Human CCAAT/enhanc
70	6	13.0	149	22	ABE17274	Novel human diagno
71	6	13.0	150	22	AAU16099	Human novel secret
72	6	13.0	152	21	AAG48562	Arabidopsis thalia
73	6	13.0	159	21	AAG08936	Arabidopsis thalia
74	6	13.0	172	22	AAU16162	Human novel secret
75	6	13.0	177	22	AAG01381	Human polypeptide
76	6	13.0	180	19	AAW59797	Human polypeptide
77	6	13.0	194	22	AAU19673	Amino acid sequenc
78	6	13.0	198	21	AAG09322	Human diagnostic a
79	6	13.0	198	21	AAG52804	Arabidopsis thalia
80	6	13.0	202	15	AAR47470	Arabidopsis thalia
81	6	13.0	212	22	ABG24830	Rat proteosome del
82	6	13.0	219	13	AAR25592	Novel human diagno
83	6	13.0	219	15	AAR47473	RING12 proteosome
84	6	13.0	219	17	AAW06438	Rat proteosome RIN

No.	Score	Window	Length	Accession	Count	Description
85	13.0	6	220	AAW98473	19	H. pylori GHPO 840
86	13.0	6	222	AAT37757	20	Chlamydia trachoma
87	13.0	6	223	AAG09321	21	Arabidopsis thalia
88	13.0	6	223	AAG52803	21	Arabidopsis thalia
89	13.0	6	224	AAB61835	22	Drosophila melanog
90	13.0	6	233	AAG09320	21	Arabidopsis thalia
91	13.0	6	233	AAG52802	21	Arabidopsis thalia
92	13.0	6	239	AAB83970	16	Human proteasome s
93	13.0	6	239	AAW06435	17	Human proteasome s
94	13.0	6	243	AAY34866	20	Chlamydia pneumoni
95	13.0	6	244	AAG81950	22	S. epidermidis ope
96	13.0	6	247	AAB63519	22	Human gastric canc
97	13.0	6	252	ABG14357	22	Novel human diagno
98	13.0	6	254	ABG12787	22	Novel human diagno
99	13.0	6	254	ABG15795	22	Novel human diagno
100	13.0	6	260	AAB96770	22	Putative P. abyssi
101	13.0	6	269	AAR46337	15	PB92 serine protea
102	13.0	6	269	AAR46338	15	Subtilisin 309 Ser
103	13.0	6	269	AAR46339	15	PB92 serine protea
104	13.0	6	269	AAR46340	15	Subtilisin 309 Ser
105	13.0	6	269	AAR46341	15	PB92 serine protea
106	13.0	6	269	AAR46342	15	Subtilisin 309 Ser
107	13.0	6	269	AAR46343	15	PB92 serine protea
108	13.0	6	269	AAR46344	15	Subtilisin 309 Ser
109	13.0	6	270	AAR27641	13	Human calcium chan
110	13.0	6	273	AAW20926	18	H. Pylori cytoplas
111	13.0	6	276	ABB69618	22	Drosophila melanog
112	13.0	6	277	AAM25545	22	Human protein sequ
113	13.0	6	279	AAB16367	22	Human nervous syst
114	13.0	6	279	AAU16524	22	Human novel secret
115	13.0	6	282	AAU05262	22	Chicken stem cell
116	13.0	6	282	AAB98364	22	Chicken SCF protei
117	13.0	6	282	AAU02489	22	Chicken SCF (stem
118	13.0	6	282	AAU02775	22	Chicken SCF (stem
119	13.0	6	282	AAB73576	22	Chicken SCF (stem
120	13.0	6	282	AAB96949	22	Chicken stem cell
121	13.0	6	283	AAW20637	18	H. pylori flagella
122	13.0	6	286	AAU32713	22	Novel human secret
123	13.0	6	289	ABB60987	22	Drosophila melanog
124	13.0	6	289	ABC23002	22	Novel human diagno
125	13.0	6	296	AAE11947	22	Mouse CCAAT/enhanc
126	13.0	6	297	AAE11948	22	Rat CCAAT/enhancer
127	13.0	6	306	AAB96808	22	Putative P. abyssi
128	13.0	6	309	AAB96496	22	Putative P. abyssi
129	13.0	6	312	ABG08404	22	Novel human diagno
130	13.0	6	317	ABB69511	22	Drosophila melanog
131	13.0	6	321	AAU34275	22	Staphylococcus aur
132	13.0	6	321	AAU37101	22	Staphylococcus aur
133	13.0	6	321	AAU37128	22	Staphylococcus aur
134	13.0	6	322	AAB19116	21	Homologue of polyp
135	13.0	6	322	AAE11945	22	Human CCAAT/enhanc
136	13.0	6	329	AAB35549	22	Haemophilus influe
137	13.0	6	329	ABG11238	22	Novel human diagno
138	13.0	6	330	AAB96503	22	Putative P. abyssi
139	13.0	6	331	AAY95781	12	5H7 single chain a
140	13.0	6	331	ABG19411	22	Novel human diagno
141	13.0	6	333	AAG72693	22	Murine OR-like pol
142	13.0	6	334	AAU67978	22	Propionibacterium
143	13.0	6	339	ABG26859	22	Novel human diagno
144	13.0	6	342	ABG08951	22	Drosophila melanog
145	13.0	6	344	AAU09816	22	Fibroblast growth
146	13.0	6	344	ABG03199	22	Novel human diagno
147	13.0	6	344	AAG81114	22	Mycobacterium tube
148	13.0	6	345	AAR14408	22	Nuclear factor C/E
149	13.0	6	345	AAE11944	22	Human CCAAT/enhanc
150	13.0	6	345	AAU09068	22	Human transcriptio
151	13.0	6	350	AAM52302	22	ActA protein fragm
152	13.0	6	350	AAB85789	22	Human kinase PKIN-
153	13.0	6	351	AAW20341	18	H. pylori flagella
154	13.0	6	355	AAW20737	18	H. pylori flagella
155	13.0	6	356	AAW39772	18	Human polypeptide
156	13.0	6	359	ABB70995	22	Drosophila melanog
157	13.0	6	360	AAY34924	20	Chlamydia pneumoni

No.	Score	Window	Length	Count	Accession	Description
158	13.0	6	360	21	AAW69364	Amino acid sequenc
159	13.0	6	360	22	ABB70816	Drosophila melanog
160	13.0	6	360	22	ABG14697	Novel human diagno
161	13.0	6	360	22	AAM41558	Human polypeptide
162	13.0	6	360	22	AAG82528	S. epidermidis ope
163	13.0	6	365	22	AAM25511	Human protein sequ
164	13.0	6	368	22	AAE01696	Human gene 1 encod
165	13.0	6	371	22	AAM40429	Human polypeptide
166	13.0	6	372	22	AAU48589	Propionibacterium
167	13.0	6	376	22	AAM52301	ActA protein fragm
168	13.0	6	376	19	AAG82506	S. epidermidis ope
169	13.0	6	377	22	AAW63692	Human secreted pro
170	13.0	6	377	22	AAM93203	Human polypeptide,
171	13.0	6	377	22	AAM38833	Human polypeptide
172	13.0	6	377	22	AAU01242	Human protein diff
173	13.0	6	379	13	AAR29630	APase protein. Fu
174	13.0	6	382	21	AAG45967	Arabidopsis thalia
175	13.0	6	384	22	AAM24101	Human EST encoded
176	13.0	6	385	21	AAG45966	Arabidopsis thalia
177	13.0	6	385	22	AAM40619	Human polypeptide
178	13.0	6	387	22	AAG45965	Arabidopsis thalia
179	13.0	6	392	22	ABB58237	Drosophila melanog
180	13.0	6	396	22	AAE04799	Neurospora crassa
181	13.0	6	397	18	AAW31628	Aspergillus oryzae
182	13.0	6	397	22	AAE04886	Human protease pro
183	13.0	6	398	16	AAR75299	Aspergillus niger
184	13.0	6	398	22	ABG10841	Novel human diagno
185	13.0	6	404	22	ABB64873	Drosophila melanog
186	13.0	6	404	22	AAG75543	Human colon cancer
187	13.0	6	410	22	AAE01672	Human gene 1 encod
188	13.0	6	412	16	AAR74207	Human death associ
189	13.0	6	412	19	AAW71369	Death associated p
190	13.0	6	412	20	AAY06478	Human tumour-assoc
191	13.0	6	412	21	AAG48540	Arabidopsis thalia
192	13.0	6	412	21	AAY93685	Amino acid sequenc
193	13.0	6	414	21	AAG48539	Arabidopsis thalia
194	13.0	6	423	21	ABB62432	Drosophila melanog
195	13.0	6	439	22	AAB19115	Polypeptide isolat
196	13.0	6	442	19	AAW98707	H. pylori GHPO 728
197	13.0	6	442	20	AAU35813	Helicobacter pylor
198	13.0	6	446	20	AAY30792	Mature secreted hum
199	13.0	6	447	21	AAG29728	Arabidopsis thalia
200	13.0	6	447	21	AAG48538	Arabidopsis thalia
201	13.0	6	449	22	AAG30230	Arabidopsis thalia
202	13.0	6	452	20	AAY00864	Cellobiohydrolase
203	13.0	6	452	20	AAW95029	A. niger cellobioh
204	13.0	6	452	22	AAG14049	Arabidopsis thalia
205	13.0	6	452	21	AAG51985	Arabidopsis thalia
206	13.0	6	455	21	AAG17438	Arabidopsis thalia
207	13.0	6	455	21	AAG21851	Arabidopsis thalia
208	13.0	6	455	22	AAG90723	C glutamicum prote
209	13.0	6	456	20	AAY17067	Human 3-OST-4 prot
210	13.0	6	459	22	AAU02077	Synthetic multi-ta
211	13.0	6	463	18	AAW55714	H. pylori ORF 09cp
212	13.0	6	465	20	ABG03158	Immature human HKN
213	13.0	6	465	22	AAY30793	Novel human diagno
214	13.0	6	466	21	AAG17437	Arabidopsis thalia
215	13.0	6	466	21	AAG21850	Arabidopsis thalia
216	13.0	6	468	21	AAB19042	Amino acid sequenc
217	13.0	6	468	21	AAG29727	Drosophila melanog
218	13.0	6	472	22	AAU09810	Putative P. abyssi
219	13.0	6	473	21	AAG51984	Human fibroblast g
220	13.0	6	474	21	AAG14048	Arabidopsis thalia
221	13.0	6	476	13	AAR20540	Arabidopsis thalia
222	13.0	6	476	20	AAY30786	Alkaline protease.
223	13.0	6	477	20	AAW97105	Protein encoded by
224	13.0	6	477	22	AAB94726	Thermostable facto
225	13.0	6	478	22	ABB58996	Human protein sequ
226	13.0	6	481	22	AAB96058	Drosophila melanog
227	13.0	6	489	13	AAR27647	Putative P. abyssi
228	13.0	6	489	22	AAG17436	Human calcium chan
229	13.0	6	493	22	AAE07168	Arabidopsis thalia
230	13.0	6	495	20	AAY30785	Mycobacterium tube

No.			Len	Ct	Accession	Description
231	6	13.0	496	20	AAY37175	Amino acid sequenc
232	6	13.0	496	21	AAG80007	D. melanogaster li
233	6	13.0	497	21	AAG21849	Arabidopsis thalia
234	6	13.0	499	13	AAR27574	The ABF-B from A.
235	6	13.0	501	22	AAB64937	Drosophila melanog
236	6	13.0	504	21	AAB24066	Human PRO943 prote
237	6	13.0	504	21	AAY92864	Human fibroblast g
238	6	13.0	504	22	AAB65179	Membrane-bound pro
239	6	13.0	504	22	AAB66264	Human PRO943 (UNO4
240	6	13.0	513	22	AAM41685	Human MANGO 003 SE
241	6	13.0	513	13	AAR27644	Human polypeptide
242	6	13.0	518	22	ABB58656	Human calcium chan
243	6	13.0	526	22	AAB19114	Drosophila melanog
244	6	13.0	529	22	AAU09808	Polypeptide isolat
245	6	13.0	529	20	AAW90018	Murine fibroblast
246	6	13.0	530	18	AAW25769	Expressed antigen
247	6	13.0	534	18	AAW10026	Human MLN 51. Hom
248	6	13.0	538	18	AAR67385	Partial sporozoite
249	6	13.0	547	16	AAY06927	Mitochondrial prot
250	6	13.0	548	20	ABB61472	C. albicans antige
251	6	13.0	553	22	AAR22925	Drosophila melanog
252	6	13.0	555	13	AAY91945	Truncated hamster
253	6	13.0	559	21	AAB93115	Human chaperone pr
254	6	13.0	559	22	AAW98592	Human protein sequ
255	6	13.0	560	19	AAU34618	H. pylori GHPO 97
256	6	13.0	561	22	AAG30029	E. coli cellular p
257	6	13.0	567	21	AAR36220	Arabidopsis thalia
258	6	13.0	571	11	AAR04713	Pseudomonas aerugi
259	6	13.0	573	18	AAW14946	Amino acid sequenc
260	6	13.0	573	18	AAW12345	Human heat shock p
261	6	13.0	573	18	AAW01657	Human heat-shock p
262	6	13.0	573	19	AAW56120	Human heat shock p
263	6	13.0	573	20	AAY23926	Protein sequence o
264	6	13.0	573	21	AAY93333	Amino acid sequenc
265	6	13.0	573	21	AAY93334	Amino acid sequenc
266	6	13.0	573	22	AAB50262	Heat shock protein
267	6	13.0	573	16	AAE11756	Human heat shock p
268	6	13.0	575	22	AAR64763	Human P1 protein.
269	6	13.0	580	22	AAB65504	Drosophila melanog
270	6	13.0	588	22	ABG12546	Novel human diagno
271	6	13.0	594	22	AAU09817	Murine FGFR-L extr
272	6	13.0	595	22	AAU38334	Salmonella typhi c
273	6	13.0	595	22	ABB32176	Peptide #4827 enco
274	6	13.0	595	22	ABB22720	Protein #4719 enco
275	6	13.0	597	21	AAY90257	Streptococcus equi
276	6	13.0	600	21	AAG40030	Zea mays protein f
277	6	13.0	608	22	AAM38643	Human polypeptide
278	6	13.0	610	22	AAM52300	ActA protein. Lis
279	6	13.0	611	13	AAR23151	Mutant thermostabl
280	6	13.0	611	21	AAB32732	Eucalyptus grandis
281	6	13.0	611	22	AAB71064	Drosophila melanog
282	6	13.0	613	22	ABG27790	Novel human diagno
283	6	13.0	615	21	AAG30228	Arabidopsis thalia
284	6	13.0	635	20	AAW75855	Human secretory pr
285	6	13.0	635	21	AAY45160	Human secreted pro
286	6	13.0	635	21	AAY71060	Human membrane tra
287	6	13.0	639	22	AAY79171	Amino acid sequenc
288	6	13.0	639	22	AAU09142	Lysteria monocytog
289	6	13.0	643	22	AAB60087	Drosophila melanog
290	6	13.0	645	21	AAY94657	Corn CCR4 transcri
291	6	13.0	683	22	ABG06295	Novel human diagno
292	6	13.0	684	13	AAR27645	Human calcium chan
293	6	13.0	692	22	ABG08784	Novel human diagno
294	6	13.0	698	21	AAY99666	Human GTPase assoc
295	6	13.0	703	15	AAR49135	Sequence of lag D
296	6	13.0	703	20	AAY30621	Amino acid sequenc
297	6	13.0	719	21	AAY31731	Bacillus sp. malto
298	6	13.0	719	21	AAY94271	Bacillus maltogeni
299	6	13.0	719	22	AAF59956	Bacillus maltogeni
300	6	13.0	719	22	AAF74216	Bacillus maltogeni
301	6	13.0	719	22	AAF74217	Bacillus maltogeni
302	6	13.0	719	22	AAF74218	Bacillus maltogeni
304	6	13.0	719	22	AAB74219	Bacillus maltogeni
305	6	13.0	719	22	AAB74220	Bacillus maltogeni
306	6	13.0	719	22	AAB74221	Bacillus maltogeni
307	6	13.0	719	22	AAB74222	Bacillus maltogeni
308	6	13.0	719	22	AAB74223	Bacillus maltogeni
309	6	13.0	755	13	AAU34979	Mutant thermostabl
310	6	13.0	766	22	AAG91691	Enterococcus faeca
311	6	13.0	785	22	ABG01797	C glutamicum prote
312	6	13.0	796	22	ABG11797	Novel human diagno
313	6	13.0	796	22	ABB59391	Novel human diagno
314	6	13.0	809	22	ABG23149	Drosophila melanog
315	6	13.0	821	13	ABG21079	Mutant thermostabl
316	6	13.0	821	22	ABG63141	Novel human diagno
317	6	13.0	828	21	AAG45432	Drosophila melanog
318	6	13.0	844	21	AAR23147	Arabidopsis thalia
319	6	13.0	857	13	AAB23271	Mutant thermostabl
320	6	13.0	860	22	ABB59530	Balanus amphitrite
321	6	13.0	862	22	AAR23148	Drosophila melanog
322	6	13.0	875	13	ABG05821	Mutant thermostabl
323	6	13.0	882	15	AAR58609	Novel human diagno
324	6	13.0	887	20	AAY49560	Hamster HMG-CoA re
325	6	13.0	888	13	AAR21830	Human protein C pr
326	6	13.0	893	13	AAR22146	Sequence of a ther
327	6	13.0	893	20	AAW78469	Mutant thermostabl
328	6	13.0	908	22	AAM00834	Thermatoga maritim
329	6	13.0	913	19	AAW58829	Human bone marrow
330	6	13.0	913	21	AAY82595	Pseudomonas OrfX s
331	6	13.0	913	22	AAB82259	Pseudomonas alcali
332	6	13.0	917	22	ABG06296	Pseudomonas alcali
333	6	13.0	917	22	ABG08780	Novel human diagno
334	6	13.0	926	22	ABG21080	Novel human diagno
335	6	13.0	951	22	AAM79950	Novel human diagno
336	6	13.0	955	21	AAB42231	Human protein SEQ
337	6	13.0	960	21	AAB45431	Human ORFX ORF1995
338	6	13.0	962	22	ABB11976	Arabidopsis thalia
339	6	13.0	977	22	AAM51671	Human TRAP150 homo
340	6	13.0	986	21	AAG45430	Phaseolus lunatus
341	6	13.0	993	13	AAR27650	Arabidopsis thalia
342	6	13.0	1015	22	ABG08785	Human calcium chan
343	6	13.0	1053	22	AAB78966	Novel human diagno
344	6	13.0	1086	22	ABB60786	Human protein SEQ
345	6	13.0	1087	22	ABG24996	Drosophila melanog
346	6	13.0	1127	22	ABG22813	Novel human diagno
347	6	13.0	1174	22	ABB61704	Novel human diagno
348	6	13.0	1217	22	AAB33177	Drosophila melanog
349	6	13.0	1218	22	ABB63445	Novel human diagno
350	6	13.0	1222	22	ABB66096	Novel human secret
351	6	13.0	1237	22	AAB20014	Drosophila melanog
352	6	13.0	1342	22	ABB64632	Drosophila melanog
353	6	13.0	1343	22	AAM00947	Drosophila SREBP c
354	6	13.0	1359	22	AAB41785	Drosophila melanog
355	6	13.0	1371	22	ABB52745	Human bone marrow
356	6	13.0	1437	22	AAM79507	Human ORFX ORF1549
357	6	13.0	1820	22	AAU27679	Escherichia coli p
358	6	13.0	1824	21	AAY21801	Human protein SEQ
359	6	13.0	1856	20	AAY83269	Human full-length
360	6	13.0	1856	21	AAU15088	Murine CACNAIF pro
361	6	13.0	1897	22	AAY49429	B. subtilis rib op
362	6	13.0	1912	21	AAR33547	Polypeptide encode
363	6	13.0	1967	14	AAW63139	Protein encoded by
364	6	13.0	1968	19	AAY49430	Human CACNAIF spli
365	6	13.0	1977	21	ABB71277	Sequence of the al
366	6	13.0	2006	22	ABB58077	Human calcium chan
367	6	13.0	2110	22	AAR72607	Human CACNAIF long
368	6	13.0	2138	16	AAB10593	Drosophila melanog
369	6	13.0	2138	21	AAR33545	Human neuronal cal
370	6	13.0	2161	14	AAR71001	Human calcium chan
371	6	13.0	2161	16	AAR71002	Human calcium chan
372	6	13.0	2161	16	AAW63137	Sequence of the al
373	6	13.0	2161	19	AAW63149	Human neuronal cal
374	6	13.0	2161	19	AAW63137	Human neuronal cal
375	6	13.0	2161	19	AAW63149	Human calcium chan
376	6	13.0	2161	21	AAB10568	Human calcium chan

No.	Score	Cnt	Grp	Len	Accession	Description
377	13.0	6	2163	16	AAR71003	Human neuronal cal
378	13.0	6	2163	21	AAB10570	Human calcium chan
379	13.0	6	2279	22	AAM38733	Human polypeptide
380	13.0	6	2297	22	AAM40519	Human polypeptide
381	13.0	6	2322	18	AAW15566	Melanoma-associate
382	13.0	6	2344	22	AAU37120	Staphylococcus aur
383	13.0	6	2778	22	ABB58683	Drosophila melanog
384	13.0	6	26926	22	AAU05396	Human titin (conne
385	10.9	6	5	14	AAR44624	Calcitonin fragmen
386	10.9	6	7	19	AAW70511	Escherichia coli g
387	10.9	5	8	14	AAR37305	Human hepatocyte g
388	10.9	5	8	15	AAR53960	N-terminal sequenc
389	10.9	5	8	18	AAW13431	Melanoma-homing pe
390	10.9	5	8	19	AAW60292	Melanoma tumour ho
391	10.9	5	8	20	AAW93716	Mouse B16B15b mela
392	10.9	5	8	21	AAB21704	Murine melanoma ho
393	10.9	5	8	21	AAB23539	Angiogenic vascula
394	10.9	5	8	21	AAB17961	Integrin-binding p
395	10.9	5	8	21	AAE06282	Tumour homing pept
396	10.9	5	9	17	AAW94247	Human leucocyte an
397	10.9	5	9	19	AAW77801	Multi-functional c
398	10.9	5	10	19	AAW69019	Linker used in hum
399	10.9	5	10	21	AAY81276	Gly/Ser-rich linke
400	10.9	5	10	22	AAG94010	Human complementar
401	10.9	5	10	22	AAG94012	Human complementar
402	10.9	5	10	22	AAG94124	Human complementar
403	10.9	5	10	22	AAG94128	Human complementar
404	10.9	5	10	22	AAG94261	Human complementar
405	10.9	5	10	22	AAG94263	Human complementar
406	10.9	5	10	22	AAG94265	Human complementar
407	10.9	5	10	22	AAG94247	Human complementar
408	10.9	5	10	22	AAG94279	Human complementar
409	10.9	5	10	22	AAG94281	Human complementar
410	10.9	5	10	22	AAG94299	Human complementar
411	10.9	5	10	22	AAG94301	Human complementar
412	10.9	5	10	22	AAG94303	Human complementar
413	10.9	5	10	22	AAG94305	Human complementar
414	10.9	5	10	22	AAG94307	Human complementar
415	10.9	5	10	22	AAG94309	Human complementar
416	10.9	5	10	22	AAG94311	Human complementar
417	10.9	5	10	22	AAG94313	Human complementar
418	10.9	5	10	22	AAG94315	Human complementar
419	10.9	5	10	22	AAG94317	Human complementar
420	10.9	5	10	22	AAG94323	Human complementar
421	10.9	5	10	22	AAG94325	Human complementar
422	10.9	5	10	22	AAG94333	Human complementar
423	10.9	5	10	22	AAG94390	Human complementar
424	10.9	5	10	22	AAG94392	Human complementar
425	10.9	5	10	22	AAG95196	Human complementar
426	10.9	5	10	22	AAG95198	Human complementar
427	10.9	5	10	22	AAG95256	Human complementar
428	10.9	5	10	22	AAG95316	Human complementar
429	10.9	5	10	22	AAG95380	Human complementar
430	10.9	5	10	22	AAG96214	Human complementar
431	10.9	5	10	22	AAG96244	Human complementar
432	10.9	5	10	22	AAG96424	Human complementar
433	10.9	5	10	22	AAG96426	Human complementar
434	10.9	5	10	22	AAG96428	Human complementar
435	10.9	5	10	22	AAG96430	Human complementar
436	10.9	5	10	22	AAG96444	Human complementar
437	10.9	5	10	22	AAG96446	Human complementar
438	10.9	5	10	22	AAG96456	Human complementar
439	10.9	5	10	22	AAG96624	Human complementar
440	10.9	5	10	22	AAG96626	Human complementar
441	10.9	5	10	22	AAG96628	Human complementar
442	10.9	5	10	22	AAG97132	Human complementar
443	10.9	5	10	22	AAG97136	Human complementar
444	10.9	5	10	22	AAG97188	Human complementar
445	10.9	5	10	22	AAG97206	Human complementar
446	10.9	5	10	22	AAG97210	Human complementar
447	10.9	5	10	22	AAG97212	Human complementar
448	10.9	5	10	22	AAG97216	Human complementar
449	10.9	5	10	22	AAG97216	Human complementar

No.	Accession	Grp	Score	Cnt	Description
450	AAG97220	10	10.9	5	Human complementar
451	AAG97222	10	10.9	5	Human complementar
452	AAG97991	10	10.9	5	Human complementar
453	AAG97992	10	10.9	5	Human complementar
454	AAG97993	10	10.9	5	Human complementar
455	AAG97994	10	10.9	5	Human complementar
456	AAG85591	10	10.9	5	Saccharomyces cere
457	AAG85593	10	10.9	5	Saccharomyces cere
458	AAG93432	10	10.9	5	Proto-oncogene C-CB
459	AAG93433	10	10.9	5	Proto-oncogene C-CB
460	AAP60483	11	10.9	5	Epstein-Barr virus
461	AAR25060	13	10.9	5	Synthetic random c
462	AAG64070	11	10.9	5	HTalpha-L11-beta 1
463	AAG63314	12	10.9	5	Amino acid sequenc
464	AAR40791	13	10.9	5	Sequence #4 from a
465	AAY06305	14	10.9	5	Mouse pancreatic-d
466	AAM97293	14	10.9	5	Human peptide #568
467	AAM97716	14	10.9	5	Human peptide #991
468	AAW77802	15	10.9	5	Multi-functional c
469	AAW69020	15	10.9	5	Linker used in hum
470	AAW46296	15	10.9	5	Linker 2 used for
471	AAY49078	15	10.9	5	Fragment of dnaB c
472	AAY22379	15	10.9	5	TPO receptor bindi
473	AAW92605	15	10.9	5	Mouse beta-actin a
474	AAW92534	16	10.9	5	Beta-actin referen
475	AAR47026	16	10.9	5	Plasminogen activa
476	AAW48746	16	10.9	5	Human granulocytic
477	AAW81961	16	10.9	5	Rat ENDO-I protein
478	AAW81962	16	10.9	5	Rat HPR protein fr
479	AAY45049	17	10.9	5	DEVD containing pe
480	AAB60913	17	10.9	5	KIX domain binding
481	AAY71888	18	10.9	5	Linker peptide #7,
482	AAR36604	18	10.9	5	N-terminus of 37 k
483	AAW69830	19	10.9	5	Peptide fragment p
484	AAY04268	19	10.9	5	Bovine vascular sm
485	AAW81955	19	10.9	5	Human ENDO-I prote
486	AAW81957	20	10.9	5	Human ENDO-I varia
487	AAW65563	20	10.9	5	Epstein-Barr virus
488	AAE09083	20	10.9	5	Epstein-Barr virus
489	AAW34053	21	10.9	5	Human MDR1-P glyco
490	AAY95779	21	10.9	5	Linker peptide. S
491	ABB40487	21	10.9	5	Peptide #7993 enco
492	ABB24820	21	10.9	5	Protein #6819 enco
493	AAM61312	21	10.9	5	Human brain expres
494	AAM74044	21	10.9	5	Human bone marrow
495	AAM20204	21	10.9	5	Peptide #6638 enco
496	AAM34211	22	10.9	5	Protein #8248 enco
497	AAB97036	22	10.9	5	Chimeric polypepti
498	ABB30607	22	10.9	5	Peptide #3258 enco
499	ABB35771	22	10.9	5	Peptide #3277 enco
500	ABB21195	22	10.9	5	Protein #3194 enco
501	AAM56580	22	10.9	5	Human brain expres
502	AAM68960	22	10.9	5	Human bone marrow
503	AAM16787	22	10.9	5	Peptide #3221 enco
504	AAM29271	22	10.9	5	Peptide #3308 enco
505	AAM04504	22	10.9	5	Peptide #3186 enco
506	AAR65099	23	10.9	5	Random biotinylati
507	AAW46702	23	10.9	5	Biotinylation pept
508	AAY29305	23	10.9	5	Biotinylation pept
509	AAW67910	23	10.9	5	Human secreted pro
510	AAY87142	23	10.9	5	Human secreted pro
511	AAY54984	24	10.9	5	Angiogenin inhibit
512	AAE12642	24	10.9	5	Biotinylated synth
513	AAU04311	24	10.9	5	ATP-binding casset
514	AAE06119	24	10.9	5	Human gene 14 enco
515	AAE09937	24	10.9	5	Biotinylated pepti
516	AAY93255	24	10.9	5	Flexible linker us
517	ABB50758	24	10.9	5	Human secreted pro
518	ABG01293	24	10.9	5	Novel human diagno
519	ABB41868	25	10.9	5	Peptide #9374 enco
520	ABB25568	25	10.9	5	Protein #7567 enco
521	AAM62741	25	10.9	5	Human brain expres

No.	Accession	Val	10.9	5	Description
523	AAM75560	25	10.9	5	Human bone marrow
524	AAM20639	22	10.9	5	Peptide #7073 enco
525	AAM35666	22	10.9	5	Peptide #9703 enco
526	AAW67973	20	10.9	5	Fragment of human
527	AAW79384	19	10.9	5	Staphylococcus aur
528	AAY38452	20	10.9	5	Human secreted pro
529	ABB39191	22	10.9	5	Peptide #6697 enco
530	ABB24049	22	10.9	5	Protein #6048 enco
531	AAM59850	22	10.9	5	Human brain expres
532	AAM72439	22	10.9	5	Human bone marrow
533	AAM32688	22	10.9	5	Peptide #6725 enco
534	AAY42794	20	10.9	5	Streptococcus pyog
535	AAB09195	21	10.9	5	Hepatitis GB virus
536	AAO08982	22	10.9	5	Human polypeptide
537	AAP61348	7	10.9	5	Sequence of calcit
538	AAP61349	7	10.9	5	Sequence of calcit
539	ABB29408	22	10.9	5	Peptide #2059 enco
540	ABB34590	22	10.9	5	Peptide #2096 enco
541	ABB19996	22	10.9	5	Protein #1995 enco
542	AAM55379	22	10.9	5	Human brain expres
543	AAM67774	22	10.9	5	Human bone marrow
544	AAM75470	22	10.9	5	Human bone marrow
545	AAM15583	22	10.9	5	Peptide #2017 enco
546	AAM28075	22	10.9	5	Peptide #2112 enco
547	AAM03327	22	10.9	5	Peptide #2009 enco
548	AAW03618	22	10.9	5	Rat bradykinin GPR
549	AAO03546	22	10.9	5	Human polypeptide
550	AAG92939	22	10.9	5	C glutamicum prote
551	AAW15081	18	10.9	5	Lac28E monomer use
552	AAB27608	21	10.9	5	Human secreted pro
553	ABG08816	22	10.9	5	Novel human diagno
554	AAR57184	15	10.9	5	GET epitope C inse
555	AAB38405	22	10.9	5	Fragment of human
556	ABG15474	22	10.9	5	Novel human diagno
557	AAM18882	22	10.9	5	Peptide #5316 enco
558	AAR57187	15	10.9	5	GET epitope C inse
559	AAB38977	21	10.9	5	Human secreted pep
560	ABG10490	22	10.9	5	Novel human diagno
561	ABB42942	22	10.9	5	Peptide #10448 enc
562	AAM63844	22	10.9	5	Human brain expres
563	AAM76657	22	10.9	5	Human bone marrow
564	AAM36762	22	10.9	5	Peptide #10799 enc
565	AAR57185	15	10.9	5	GET epitope C inse
566	AAY60359	20	10.9	5	Human normal bladd
567	AAM60530	22	10.9	5	Human brain expres
568	AAM73186	22	10.9	5	Human bone marrow
569	AAM33396	22	10.9	5	Peptide #7433 enco
570	AAW27950	18	10.9	5	Staphylococcus aur
571	AAY54265	21	10.9	5	Amino acid fragmen
572	AAY55696	21	10.9	5	B. subtilis ycfB p
573	AAU01783	22	10.9	5	Human secreted pro
574	AAP80076	9	10.9	5	Cytomegalovirus (C
575	AAW18235	18	10.9	5	Fibrobacter succin
576	AAG55145	21	10.9	5	Arabidopsis thalia
577	AAY39971	20	10.9	5	Fusarium poae hydr
578	AAB43813	21	10.9	5	Peptide #11319 enc
579	ABB22221	22	10.9	5	Protein #4220 enco
580	ABB26750	22	10.9	5	Protein #8749 enco
581	AAM64779	22	10.9	5	Human brain expres
582	AAM77538	22	10.9	5	Human bone marrow
583	AAM21471	22	10.9	5	Peptide #7905 enco
584	AAM37722	22	10.9	5	Peptide #11759 enc
585	AAU01625	22	10.9	5	Human secreted pro
586	AAR57186	15	10.9	5	GET epitope C inse
587	AAB50421	21	10.9	5	Human uncoupling p
588	AAB27164	21	10.9	5	HPIV2 partial prot
589	ABB31202	22	10.9	5	Peptide #3853 enco
590	ABB36400	22	10.9	5	Protein #3906 enco
591	AAM57165	22	10.9	5	Human brain expres
592	AAM69564	22	10.9	5	Human bone marrow
593	AAM17385	22	10.9	5	Peptide #3819 enco
594	AAM29904	22	10.9	5	Peptide #3941 enco
595	AAM05074	22	10.9	5	Peptide #3756 enco

No.	Accession	Val1	Val2	10.9	5	Description
596	ABG12103	22	42	10.9	5	Novel human diagno
597	ABG12668	22	42	10.9	5	Novel human diagno
598	AAB67213	22	42	10.9	5	HER2 peptide ligan
599	AAY60379	20	43	10.9	5	Human normal bladd
600	AAM86005	22	43	10.9	5	Human immune/haema
601	AAB82905	21	43	10.9	5	Human TNF receptor
602	AAB27604	22	44	10.9	5	Human secreted pro
603	ABG12084	22	44	10.9	5	Novel human diagno
604	ABG28609	22	44	10.9	5	Novel human diagno
605	AAO00262	22	44	10.9	5	Human polypeptide
606	AAW88773	20	45	10.9	5	Polypeptide fragme
607	AAB63064	21	45	10.9	5	Human secreted pro
608	ABB50606	22	45	10.9	5	Human secreted pro
609	AAM82748	22	45	10.9	5	Human immune/haema
610	AAM24420	18	46	10.9	5	Nucleic acid deliv
611	AAY30849	20	46	10.9	5	Human secreted pro
612	AAM89492	22	46	10.9	5	Human immune/haema
613	AAO09653	22	46	10.9	5	Human polypeptide
614	AAG76210	22	46	10.9	5	Human colon cancer
615	ABB41431	22	47	10.9	5	Peptide #8937 enco
616	AAM62304	22	47	10.9	5	Human brain expres
617	AAM75107	22	47	10.9	5	Human bone marrow
618	AAM35223	22	47	10.9	5	Peptide #9260 enco
619	AAW27749	18	48	10.9	5	Amino acid sequenc
620	AAO09043	22	48	10.9	5	Human polypeptide
621	AAO09243	22	48	10.9	5	Human polypeptide
622	AAO11779	22	48	10.9	5	Human polypeptide
623	AAG76454	22	48	10.9	5	Human colon cancer
624	AAR65749	16	49	10.9	5	Linker sequence us
625	AAW98307	19	49	10.9	5	H. pylori GHPO 535
626	AAB27809	21	49	10.9	5	Human secreted pro
627	AAG37929	21	49	10.9	5	Arabidopsis thalia
628	AAE14027	22	49	10.9	5	Chemically modifie
629	ABB42096	22	49	10.9	5	Peptide #9602 enco
630	ABB14832	22	49	10.9	5	Human nervous syst
631	AAU22665	22	49	10.9	5	Novel human colon
632	AAM62976	22	49	10.9	5	Human brain expres
633	AAM75789	22	49	10.9	5	Human bone marrow
634	AAM82540	22	49	10.9	5	Human immune/haema
635	AAM92671	22	49	10.9	5	Human digestive sy
636	AAO11940	22	49	10.9	5	Human polypeptide
637	AAM35899	22	49	10.9	5	Peptide #9936 enco
638	AAW28348	18	50	10.9	5	Staphylococcus aur
639	AAG37777	21	50	10.9	5	Arabidopsis thalia
640	AAU50306	22	50	10.9	5	Propionibacterium
641	AAU59594	22	50	10.9	5	Propionibacterium
642	ABB23488	22	50	10.9	5	Protein #5487 enco
643	AAM58929	22	51	10.9	5	Human brain expres
644	AAR94977	17	51	10.9	5	Calcium channel is
645	AAW32466	18	51	10.9	5	Mycobacterium tube
646	AAW32339	18	51	10.9	5	Mycobacterium tube
647	AAW81701	19	51	10.9	5	M. tuberculosis im
648	AAW64334	19	51	10.9	5	Mycobacterium tube
649	AAY39131	20	51	10.9	5	M. tuberculosis ES
650	AAY38988	20	51	10.9	5	M. tuberculosis re
651	AAY38451	20	51	10.9	5	Human secreted pro
652	AAG34437	21	51	10.9	5	Arabidopsis thalia
653	ABB10901	22	51	10.9	5	Human ovarian and/
654	AAM96224	22	51	10.9	5	Human reproductive
655	AAU01897	22	51	10.9	5	Mycobacterium tube
656	AAG00521	22	52	10.9	5	Human secreted pro
657	AAU43770	22	52	10.9	5	Propionibacterium
658	ABB16203	22	52	10.9	5	Human nervous syst
659	AAM06584	22	52	10.9	5	Human foetal prote
660	AAM06729	22	53	10.9	5	Human foetal prote
661	AAY42792	20	53	10.9	5	Streptococcus pyog
662	AAB38489	21	54	10.9	5	Human secreted pro
663	AAU43821	22	54	10.9	5	Propionibacterium
664	ABG14477	22	54	10.9	5	Novel human diagno
665	ABB40347	22	54	10.9	5	Peptide #7853 enco
666	ABB17231	22	54	10.9	5	Human nervous syst
667	ABB24735	22	54	10.9	5	Protein #6734 enco
668	AAM81171	22	54	10.9	5	Human haematologic

No.	n	Tm	Len	k	Accession	Description
669	5	10.9	54	22	AAM81182	Human haematologic
670	5	10.9	54	22	AAM81581	Human haematologic
671	5	10.9	54	22	AAM61153	Human brain expres
672	5	10.9	54	22	AAM73867	Human bone marrow
673	5	10.9	54	22	AAM34049	Peptide #8086 enco
674	5	10.9	54	22	AAU40523	Peptide #8086 enco
675	5	10.9	55	22	ABB29477	Propionibacterium
676	5	10.9	55	22	ABB31000	Peptide #2128 enco
677	5	10.9	55	22	AAM34655	Peptide #3651 enco
678	5	10.9	55	22	ABB36193	Peptide #2161 enco
679	5	10.9	55	22	ABB15421	Human polypeptide
680	5	10.9	55	22	ABB20062	Peptide #3699 enco
681	5	10.9	55	22	AAM53445	Human nervous syst
682	5	10.9	55	22	AAM56974	Protein #2061 enco
683	5	10.9	55	22	AAM60182	Human brain expres
684	5	10.9	55	22	AAM67836	Human brain expres
685	5	10.9	55	22	AAM69361	Human bone marrow
686	5	10.9	55	22	AAM72800	Human bone marrow
687	5	10.9	55	22	AAO10860	Human bone marrow
688	5	10.9	55	22	AAM15651	Human polypeptide
689	5	10.9	55	22	AAM28152	Peptide #2085 enco
690	5	10.9	55	22	AAM33030	Peptide #2189 enco
691	5	10.9	55	22	AAM03390	Peptide #7067 enco
692	5	10.9	55	22	AAM04888	Peptide #2072 enco
693	5	10.9	55	22	AAW01174	Peptide #3570 enco
694	5	10.9	56	17	AAG16991	Serine protease Pf
695	5	10.9	56	21	AAG48841	Arabidopsis thalia
696	5	10.9	56	21	AAU60540	Arabidopsis thalia
697	5	10.9	56	22	ABB03535	Propionibacterium
698	5	10.9	56	22	AAB50570	Human musculoskele
699	5	10.9	56	22	AAB38625	Flea serine protea
700	5	10.9	57	21	AAM84903	Human secreted pro
701	5	10.9	57	22	AAM23726	Human immune/haema
702	5	10.9	57	22	AAM24492	Human EST encoded
703	5	10.9	57	21	AAG51395	Human EST encoded
704	5	10.9	58	22	AAC76309	Arabidopsis thalia
705	5	10.9	58	22	AAM84368	Human colon cancer
706	5	10.9	59	22	AAO13672	Human immune/haema
707	5	10.9	59	22	AAR65750	Human polypeptide
708	5	10.9	60	16	AAW27837	Linker sequence us
709	5	10.9	60	18	AAY03211	Staphylococcus aur
710	5	10.9	60	20	AAG02747	Amino acid sequenc
711	5	10.9	60	21	AAE14028	Human secreted pro
712	5	10.9	60	22	AAU55956	Chemically modifie
713	5	10.9	60	22	AAM84698	Propionibacterium
714	5	10.9	61	22	AAU47353	Human immune/haema
715	5	10.9	61	22	ABG04669	Propionibacterium
716	5	10.9	61	22	ABG09797	Novel human diagno
717	5	10.9	61	22	AAM83024	Novel human diagno
718	5	10.9	61	22	AAO12505	Human immune/haema
719	5	10.9	61	21	AAM34466	Human polypeptide
720	5	10.9	62	21	AAG33304	Human secreted pro
721	5	10.9	62	22	AAG91645	Zea mays protein f
722	5	10.9	62	19	AAY20951	C glutamicum prote
723	5	10.9	63	21	AAG18764	Human presenilin I
724	5	10.9	63	21	AAM32669	Zea mays protein f
725	5	10.9	63	21	AAG44401	Zea mays protein f
726	5	10.9	63	22	AAM81044	Arabidopsis thalia
727	5	10.9	63	22	AAM82041	Human haematologic
728	5	10.9	63	22	AAM95167	Human haematologic
729	5	10.9	63	22	AAM83389	Human reproductive
730	5	10.9	63	22	AAM84472	Human immune/haema
731	5	10.9	63	22	AAO09663	Human immune/haema
732	5	10.9	63	22	AAU20820	Human immune/haema
733	5	10.9	64	22	AAM23862	Propionibacterium
734	5	10.9	64	22	AAB85559	Human polypeptide
735	5	10.9	64	22	AAW19268	Human novel foetal
736	5	10.9	64	18	AAY48384	Human EST encoded
737	5	10.9	65	20	AAY12493	Human secreted pro
738	5	10.9	65	20	AAU58313	Lactobacillus amyl
739	5	10.9	65	20	ABB38959	Human prostate can
740	5	10.9	65	22		Human 5' EST secre
741	5	10.9	65	22		Propionibacterium
						Peptide #6465 enco
742	5	10.9	65	22	ABB39031	Peptide #6537 enco
743	5	10.9	65	22	ABB23975	Protein #5974 enco
744	5	10.9	65	22	AAM59610	Human brain expres
745	5	10.9	65	22	AAM59685	Human brain expres
746	5	10.9	65	22	AAM72183	Human bone marrow
747	5	10.9	65	22	AAM72261	Human bone marrow
748	5	10.9	65	22	AAM19520	Peptide #5954 enco
749	5	10.9	65	22	AAM32443	Peptide #6480 enco
750	5	10.9	65	22	AAM32520	Peptide #6557 enco
751	5	10.9	65	22	AAM41112	Human polypeptide
752	5	10.9	65	22	AAM06614	Human foetal prote
753	5	10.9	66	22	ABB44134	Peptide #11640 enc
754	5	10.9	66	22	ABB23022	Protein #5021 enco
755	5	10.9	66	22	ABB27016	Protein #9015 enco
756	5	10.9	66	22	AAM65157	Human brain expres
757	5	10.9	66	22	AAM77862	Human bone marrow
758	5	10.9	66	22	AAM21762	Peptide #8196 enco
759	5	10.9	66	22	AAM38084	Peptide #12121 enc
760	5	10.9	66	22	AAB85512	Human protein kina
761	5	10.9	66	22	ABB27810	Human peptide #461
762	5	10.9	67	22	AAM32981	Peptide #487 encod
763	5	10.9	67	22	ABB18455	Protein #454 encod
764	5	10.9	67	22	AAO22314	Human cardiovascul
765	5	10.9	67	22	AAM94971	Human reproductive
766	5	10.9	67	22	AAM53779	Human brain expres
767	5	10.9	67	22	AAM66165	Human bone marrow
768	5	10.9	67	22	AAO11655	Human polypeptide
769	5	10.9	67	22	AAO12955	Human polypeptide
770	5	10.9	67	22	AAM14036	Peptide #470 encod
771	5	10.9	67	22	AAM26442	Peptide #479 encod
772	5	10.9	67	22	AAM01777	Peptide #459 encod
773	5	10.9	68	20	AAY07905	Human secreted pro
774	5	10.9	68	22	ABG00305	Novel human diagno
775	5	10.9	68	22	AAM88231	Human immune/haema
776	5	10.9	68	22	AAM88461	Human immune/haema
777	5	10.9	68	22	AAO13010	Human polypeptide
778	5	10.9	69	22	ABB30649	Peptide #300 enco
779	5	10.9	69	22	ABB30747	Peptide #3398 enco
780	5	10.9	69	22	ABB35820	Peptide #3326 enco
781	5	10.9	69	22	AAM35924	Peptide #3430 enco
782	5	10.9	69	22	ABB21236	Peptide #3235 enco
783	5	10.9	69	22	ABB21333	Protein #3332 enco
784	5	10.9	69	22	AAM56727	Human brain expres
785	5	10.9	69	22	AAM69004	Human bone marrow
786	5	10.9	69	22	AAM69105	Human immune/haema
787	5	10.9	69	22	AAM90904	Human immune/haema
788	5	10.9	69	22	AAM16941	Peptide #3375 enco
789	5	10.9	69	22	AAM29321	Peptide #3358 enco
790	5	10.9	69	22	AAM29426	Peptide #3463 enco
791	5	10.9	69	22	AAM04634	Peptide #316 enco
792	5	10.9	70	22	AAR97850	Rat brain calcium
793	5	10.9	70	17	AAB56896	Human prostate can
794	5	10.9	70	22	AAU50030	Propionibacterium
795	5	10.9	70	22	AAU52829	Propionibacterium
796	5	10.9	70	22	AAU59695	Propionibacterium
797	5	10.9	71	20	AAY11890	Human 5' EST secre
798	5	10.9	71	21	AAG33189	Pinus radiata tran
799	5	10.9	71	21	AAG18763	Zea mays protein f
800	5	10.9	71	21	AAG32668	Zea mays protein f
801	5	10.9	71	21	AAG37983	Arabidopsis thalia
802	5	10.9	71	21	AAG48400	Arabidopsis thalia
803	5	10.9	71	22	ABB04010	Human musculoskele
804	5	10.9	71	21	AAM83013	Human immune/haema
805	5	10.9	71	22	AAM88461	Human immune/haema
806	5	10.9	72	22	AAB15198	Human nervous syst
807	5	10.9	72	22	AAU29770	Novel human secret
808	5	10.9	72	22	AAM96288	Human reproductive
809	5	10.9	72	22	AAM42317	Human breast or ov
810	5	10.9	73	20	AAY73882	Human prostate tum
811	5	10.9	73	21	ABG227287	Novel human diagno
812	5	10.9	74	20	AAU56270	Propionibacterium
813	5	10.9	74	22	AAM96322	Human reproductive
814	5	10.9	74	22	AAM39326	Human polypeptide

No.					Accession	Description
815	5	10.9	74	22	AAB64402	Amino acid sequenc
816	5	10.9	75	21	AAB08492	Amino acid sequenc
817	5	10.9	75	21	AAG20590	Arabidopsis thalia
818	5	10.9	75	22	AAG67855	Murine leukemia vi
819	5	10.9	75	22	AAB87445	Human gene 42 enco
820	5	10.9	76	18	AAW20311	H. pylori secreted
821	5	10.9	76	22	AAB45107	Human secreted pro
822	5	10.9	76	21	ABB14839	Human nervous syst
823	5	10.9	76	22	AAM61365	Human brain expres
824	5	10.9	76	22	AAM74110	Human bone marrow
825	5	10.9	76	22	AAM84253	Human immune/haema
826	5	10.9	76	22	AAM86311	Human immune/haema
827	5	10.9	76	22	AAM93165	Human digestive sy
828	5	10.9	76	22	AAO01128	Human polypeptide
829	5	10.9	76	22	AAO07910	H. pylori secreted
830	5	10.9	77	18	AAW20663	Human secreted pro
831	5	10.9	77	19	AAW75001	Human polypeptide
832	5	10.9	77	21	AAW60998	Streptococcus pneu
833	5	10.9	77	21	AAB34363	Human secreted pro
834	5	10.9	77	22	ABB28103	Human peptide #754
835	5	10.9	77	22	ABB33278	Peptide #784 encod
836	5	10.9	77	22	ABB18740	Protein #739 encod
837	5	10.9	77	22	AAM54070	Human brain expres
838	5	10.9	77	22	AAM66458	Human bone marrow
839	5	10.9	77	22	AAM70377	Human bone marrow
840	5	10.9	77	22	AAM14330	Peptide #764 encod
841	5	10.9	77	22	AAM26740	Peptide #777 encod
842	5	10.9	77	22	AAM02059	Peptide #741 encod
843	5	10.9	78	18	AAW23001	Canine herpesvirus
844	5	10.9	78	19	AAW72661	Canine herpes viru
845	5	10.9	78	20	AAY08239	Human cadherin-3 p
846	5	10.9	78	21	AAG54871	Arabidopsis thalia
847	5	10.9	78	22	AAM8026	Human haematologic
848	5	10.9	78	22	AAM64235	Human brain expres
849	5	10.9	78	22	AAB51318	Canine herpes viru
850	5	10.9	79	22	AAB63473	Human gastric canc
851	5	10.9	79	22	AAB43069	Human ORFX ORF2833
852	5	10.9	79	21	AAG04436	Staphylococcus aur
853	5	10.9	79	21	AAG09640	Arabidopsis thalia
854	5	10.9	79	21	AAG42421	Arabidopsis thalia
855	5	10.9	79	21	AAG49868	Arabidopsis thalia
856	5	10.9	79	21	AAG03620	Human secreted pro
857	5	10.9	79	22	AAU32616	Novel human secret
858	5	10.9	79	22	AAO03643	Human polypeptide
859	5	10.9	80	20	AAW89911	Antigen 2 from clu
860	5	10.9	80	21	AAG02452	Human secreted pro
861	5	10.9	80	22	ABB69642	Drosophila melanog
862	5	10.9	80	22	AAG78902	Human 2-hydroxy ac
863	5	10.9	80	22	AAU33267	Novel human enzyme
864	5	10.9	80	22	AAO10298	Human polypeptide
865	5	10.9	80	22	AAB63490	Human gastric canc
866	5	10.9	81	22	AAU42097	Propionibacterium
867	5	10.9	81	22	AAU67375	Propionibacterium
868	5	10.9	82	13	AAR27371	TRFP Chain #2 C2ST
869	5	10.9	82	22	AAG27872	Arabidopsis thalia
870	5	10.9	82	22	AAU32043	Novel human secret
871	5	10.9	82	22	AAO05273	Human polypeptide
872	5	10.9	82	22	AAM23486	Human EST encoded
873	5	10.9	82	18	AAW27792	Human polypeptide
874	5	10.9	83	22	AAB40416	Human gastric canc
875	5	10.9	83	21	AAB24737	Human ORFX ORF180
876	5	10.9	83	21	AAG51439	Plant SDF encoded
877	5	10.9	83	21	AAG51444	Arabidopsis thalia
878	5	10.9	83	21	AAG51450	Arabidopsis thalia
879	5	10.9	83	21	AAG51454	Arabidopsis thalia
880	5	10.9	83	21	AAG51457	Arabidopsis thalia
881	5	10.9	83	22	AAM89159	Human immune/haema
882	5	10.9	83	22	AAM89775	Human immune/haema
883	5	10.9	84	21	AAG06190	Arabidopsis thalia
884	5	10.9	84	21	AAG11537	Arabidopsis thalia
885	5	10.9	84	21	AAG26576	Arabidopsis thalia
886	5	10.9	84	21	AAG26603	Arabidopsis thalia
887	5	10.9	84	21	AAG27871	Arabidopsis thalia

No.					Accession	Description
888	5	10.9	84	21	AAG35114	Arabidopsis thalia
889	5	10.9	84	21	AAG36650	Arabidopsis thalia
890	5	10.9	84	21	AAG43509	Arabidopsis thalia
891	5	10.9	84	21	AAG44506	Arabidopsis thalia
892	5	10.9	84	21	AAG46379	Arabidopsis thalia
893	5	10.9	84	21	AAG46391	Arabidopsis thalia
894	5	10.9	84	21	AAG46415	Arabidopsis thalia
895	5	10.9	84	21	AAG50740	Arabidopsis thalia
896	5	10.9	84	21	AAG50749	Human secreted pro
897	5	10.9	84	21	AAG02176	Human secreted pro
898	5	10.9	84	22	AAM89076	Human immune/haema
899	5	10.9	84	22	AAM06597	Human foetal prote
900	5	10.9	84	22	AAM06793	Human foetal prote
901	5	10.9	84	22	AAB63486	Human gastric canc
902	5	10.9	85	21	AAG37314	Arabidopsis thalia
903	5	10.9	85	21	AAG48399	Arabidopsis thalia
904	5	10.9	85	22	AAO07257	Human polypeptide
905	5	10.9	85	22	AAB63482	Human gastric canc
906	5	10.9	86	21	AAB50420	Human uncoupling p
907	5	10.9	86	21	AAG15114	Arabidopsis thalia
908	5	10.9	86	22	AAU12401	Propionibacterium
909	5	10.9	86	22	AAU62283	Propionibacterium
910	5	10.9	86	22	AAG08617	Novel human diagno
911	5	10.9	86	22	ABB50756	Human secreted pro
912	5	10.9	86	22	ABB11368	Human secreted pro
913	5	10.9	87	18	AAW18330	Bovine p32 ll-cis-
914	5	10.9	87	20	AAY35743	Chlamydia pneumoni
915	5	10.9	87	20	AAY01996	PCR clone 194, a f
916	5	10.9	87	21	AAG09639	Arabidopsis thalia
917	5	10.9	87	21	AAG20589	Arabidopsis thalia
918	5	10.9	87	21	AAG42420	Arabidopsis thalia
919	5	10.9	87	21	AAG49867	Arabidopsis thalia
920	5	10.9	87	21	AAG54324	Arabidopsis thalia
921	5	10.9	87	21	AAG56339	Arabidopsis thalia
922	5	10.9	87	22	AAM82901	Human immune/haema
923	5	10.9	87	22	AAG92990	C glutamicum prote
924	5	10.9	88	17	AAW01206	Serine protease Pf
925	5	10.9	88	19	AAW77527	Staphylococcus aur
926	5	10.9	88	20	AAY73898	Human prostate tum
927	5	10.9	88	21	AAG37313	Arabidopsis thalia
928	5	10.9	88	21	AAG41556	Arabidopsis thalia
929	5	10.9	88	22	AAU21374	Human novel foetal
930	5	10.9	88	22	AAM82687	Human immune/haema
931	5	10.9	88	22	AAM88601	Human immune/haema
932	5	10.9	88	22	AAM06338	Human foetal prote
933	5	10.9	88	22	AAB62383	Shiga toxin beta s
934	5	10.9	88	22	AAB61901	Mouse apoptosis me
935	5	10.9	88	22	AAB50603	Flea serine protea
936	5	10.9	89	13	AAR22429	Murine CP-10. Mus
937	5	10.9	89	17	AAW06404	Verotoxigenic E. c
938	5	10.9	89	19	AAY58828	Shiga toxin type 1
939	5	10.9	89	20	AAY59968	Human endometrium
940	5	10.9	89	20	AAY29338	Human secreted pro
941	5	10.9	89	21	AAG41006	Human ORFX ORF770
942	5	10.9	89	21	AAG13423	Arabidopsis thalia
943	5	10.9	89	21	AAG36152	Arabidopsis thalia
944	5	10.9	89	21	AAY96682	E. coli verotoxin
945	5	10.9	89	22	AAB44594	Mouse wound healin
946	5	10.9	89	22	AAG39089	Human secreted pro
947	5	10.9	89	22	AAO05291	Human polypeptide
948	5	10.9	89	22	AAO11912	Human polypeptide
949	5	10.9	89	22	AAB60523	Bovine ghrelin pre
950	5	10.9	90	21	AAG42419	Arabidopsis thalia
951	5	10.9	90	21	AAG02990	Human secreted pro
952	5	10.9	90	22	AAM96654	Human reproductive
953	5	10.9	90	22	AAM84788	Human immune/haema
954	5	10.9	90	22	AAU14264	Human novel protei
955	5	10.9	90	22	AAU14500	Human novel protei
956	5	10.9	91	20	AAY14480	Fragment of human
957	5	10.9	91	21	AAG18762	Zea mays protein f
958	5	10.9	91	21	AAG37981	Arabidopsis thalia
959	5	10.9	91	21	AAG00799	Human secreted pro
960	5	10.9	91	21	AAG00800	Human secreted pro

961	5	10.9	91	22	ABG21730	Novel human diagno
962	5	10.9	91	22	ABB03894	Human musculoskele
963	5	10.9	91	22	AAM85962	Human immune/haema
964	5	10.9	91	22	AAB63492	Human gastric canc
965	5	10.9	92	15	AAR65454	T-cell receptor V-
966	5	10.9	92	20	AAY37880	Chlamydia trachoma
967	5	10.9	92	21	AAB09885	Hsp70 C-terminal 9
968	5	10.9	92	21	AAG25229	Arabidopsis thalia
969	5	10.9	92	21	AAG34610	Arabidopsis thalia
970	5	10.9	92	21	AAG01138	Human secreted pro
971	5	10.9	92	22	AAU48287	Propionibacterium
972	5	10.9	92	22	ABG08059	Novel human diagno
973	5	10.9	92	22	AAO00568	Human polypeptide
974	5	10.9	93	21	AAG24552	Arabidopsis thalia
975	5	10.9	93	21	AAG56414	Arabidopsis thalia
976	5	10.9	93	22	AAU36001	Helicobacter pylor
977	5	10.9	93	22	AAU56795	Propionibacterium
978	5	10.9	93	22	ABG19225	Novel human diagno
979	5	10.9	94	19	AAW79096	Human secreted pro
980	5	10.9	94	21	AAB57028	Human prostate can
981	5	10.9	94	21	AAG36543	Arabidopsis thalia
982	5	10.9	94	21	AAG55575	Arabidopsis thalia
983	5	10.9	94	22	ABG28213	Novel human diagno
984	5	10.9	94	22	AAU31580	Novel human secret
985	5	10.9	95	16	AAW11494	ESAT6. Mycobacter
986	5	10.9	95	18	AAW28022	Amino acid sequenc
987	5	10.9	95	18	AAW67637	MSRV-1 virus clone
988	5	10.9	95	19	AAW85049	Amino acid sequenc
989	5	10.9	95	20	AAY66143	Human bladder tumo
990	5	10.9	95	20	AAY59742	Human normal ovari
991	5	10.9	95	20	AAY29890	Mycobacterium tube
992	5	10.9	95	20	AAY29788	Mycobacterial tube
993	5	10.9	95	20	AAB44119	Human cancer assoc
994	5	10.9	95	21	AAG25735	Arabidopsis thalia
995	5	10.9	95	21	AAG37927	Arabidopsis thalia
996	5	10.9	95	21	AAG60424	Arabidopsis thalia
997	5	10.9	95	22	AAB35219	M tuberculosis Rv3
998	5	10.9	95	22	AAB40905	Human ORFX ORF669
999	5	10.9	96	22	ABG10913	Novel human diagno
1000	5	10.9	97	12	AAR12123	TRFP chain 2 - tru

ALIGNMENTS

```
RESULT 1
AAO00376
ID   AAO00376 standard; Protein; 115 AA.
XX
AC   AAO00376;
XX
DT   06-NOV-2001   (first entry)
XX
DE   Human polypeptide SEQ ID NO 14268.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI80307.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 14268; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note:The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   115 AA;

  Query Match             15.2%;  Score 7;  DB 22;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         21 IIIGGSG 27
              |||||||
Db         28 IIIGGSG 34


RESULT 2
AAO01734
ID   AAO01734 standard; Protein; 118 AA.
XX
AC   AAO01734;
XX
DT   06-NOV-2001   (first entry)
XX
DE   Human polypeptide SEQ ID NO 15626.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI81665.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
```

```
XX
PS   Claim 20; SEQ ID NO 15526; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  118 AA;

Query Match             15.2%;  Score 7;  DB 22;  Pred. No. 20;  Length 118;
Best Local Similarity  100.0%;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      6 SLKNDFK 12
          |||||||
Db    108 SLKNDFK 114

RESULT 3
AAO00422
ID   AAO00422 standard; Protein; 126 AA.
XX
AC   AAO00422;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 14314.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI80353.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 14314; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
```

```
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  126 AA;

Query Match             15.2%;  Score 7;  DB 22;  Length 126;
Best Local Similarity  100.0%;  Pred. No. 21;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     21 IIIGGSG 27
          |||||||
Db     67 IIIGGSG 73

RESULT 4
AAO11182
ID   AAO11182 standard; Protein; 150 AA.
XX
AC   AAO11182;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 25074.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI91113.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 25074; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
```

Wed Nov 13 14:19:03 2002   us-09-833-017b-2.oligo.rag

```
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   150 AA;

     Query Match      15.2%;  Score 7;  DB 22;  Length 150;
     Best Local Similarity   100.0%;  Pred. No. 25;
     Matches      7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      21 IIIGGSG 27
            |||||||
Db      68 IIIGGSG 74


RESULT 5
AAG19904
ID   AAG19904 standard; Protein; 186 AA.
XX
AC   AAG19904;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 21882.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   06-JUL-1999;  99US-0142390.
PR   08-JUL-1999;  99US-0142803.
PR   09-JUL-1999;  99US-0142920.
PR   12-JUL-1999;  99US-0142977.
PR   13-JUL-1999;  99US-0143542.
PR   14-JUL-1999;  99US-0143624.
PR   15-JUL-1999;  99US-0144005.
PR   16-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   21-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   22-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   22-JUL-1999;  99US-0145089.
PR   23-JUL-1999;  99US-0145192.
PR   23-JUL-1999;  99US-0145145.
PR   23-JUL-1999;  99US-0145218.
PR   26-JUL-1999;  99US-0145224.
PR   27-JUL-1999;  99US-0145275.
PR   27-JUL-1999;  99US-0145913.
PR   28-JUL-1999;  99US-0145918.
PR   02-AUG-1999;  99US-0145919.
PR   02-AUG-1999;  99US-0145951.
PR   03-AUG-1999;  99US-0146386.
PR   04-AUG-1999;  99US-0146388.
PR   04-AUG-1999;  99US-0146389.
PR   05-AUG-1999;  99US-0147038.
PR   05-AUG-1999;  99US-0147204.
PR   06-AUG-1999;  99US-0147302.
                   99US-0147192.
                   99US-0147260.
                   99US-0147303.
```

```
Qy   25 GSGSLST 31
        |||||||
Db   17 GSGSLST 23


RESULT 6
AAG19903
ID   AAG19903 standard; Protein; 257 AA.
XX
AC   AAG19903;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 21881.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   14-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.

PR   06-AUG-1999; 99US-0147416.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149723.
PR   20-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
PR   27-AUG-1999; 99US-0151080.
PR   30-AUG-1999; 99US-0151303.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   10-SEP-1999; 99US-0153070.
PR   13-SEP-1999; 99US-0153758.
PR   15-SEP-1999; 99US-0154018.
PR   16-SEP-1999; 99US-0154039.
PR   20-SEP-1999; 99US-0154779.
PR   22-SEP-1999; 99US-0155139.
PR   23-SEP-1999; 99US-0155486.
PR   24-SEP-1999; 99US-0155659.
PR   28-SEP-1999; 99US-0156458.
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157117.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.

Query Match    15.2%;  Score 7;  DB 21;  Length 186;
Best Local Similarity   100.0%;  Pred. No. 30;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
PR  17-JUN-1999;  99US-0139492.        PR  13-AUG-1999;  99US-0148684.
PR  18-JUN-1999;  99US-0139454.        PR  16-AUG-1999;  99US-0149368.
PR  18-JUN-1999;  99US-0139455.        PR  17-AUG-1999;  99US-0149175.
PR  18-JUN-1999;  99US-0139456.        PR  18-AUG-1999;  99US-0149426.
PR  18-JUN-1999;  99US-0139457.        PR  20-AUG-1999;  99US-0149722.
PR  18-JUN-1999;  99US-0139458.        PR  20-AUG-1999;  99US-0149723.
PR  18-JUN-1999;  99US-0139459.        PR  20-AUG-1999;  99US-0149929.
PR  18-JUN-1999;  99US-0139460.        PR  23-AUG-1999;  99US-0149902.
PR  18-JUN-1999;  99US-0139461.        PR  23-AUG-1999;  99US-0149930.
PR  18-JUN-1999;  99US-0139462.        PR  25-AUG-1999;  99US-0150566.
PR  18-JUN-1999;  99US-0139463.        PR  26-AUG-1999;  99US-0150884.
PR  18-JUN-1999;  99US-0139750.        PR  27-AUG-1999;  99US-0151065.
PR  18-JUN-1999;  99US-0139763.        PR  27-AUG-1999;  99US-0151066.
PR  21-JUN-1999;  99US-0139817.        PR  27-AUG-1999;  99US-0151080.
PR  22-JUN-1999;  99US-0139899.        PR  30-AUG-1999;  99US-0151303.
PR  23-JUN-1999;  99US-0140353.        PR  31-AUG-1999;  99US-0151438.
PR  23-JUN-1999;  99US-0140354.        PR  01-SEP-1999;  99US-0151930.
PR  24-JUN-1999;  99US-0140695.        PR  07-SEP-1999;  99US-0152363.
PR  28-JUN-1999;  99US-0140823.        PR  10-SEP-1999;  99US-0153070.
PR  29-JUN-1999;  99US-0140991.        PR  13-SEP-1999;  99US-0153758.
PR  30-JUN-1999;  99US-0141287.        PR  15-SEP-1999;  99US-0154018.
PR  01-JUL-1999;  99US-0141842.        PR  16-SEP-1999;  99US-0154039.
PR  01-JUL-1999;  99US-0142154.        PR  20-SEP-1999;  99US-0154779.
PR  02-JUL-1999;  99US-0142055.        PR  22-SEP-1999;  99US-0155139.
PR  06-JUL-1999;  99US-0142390.        PR  23-SEP-1999;  99US-0155486.
PR  08-JUL-1999;  99US-0142803.        PR  24-SEP-1999;  99US-0155659.
PR  09-JUL-1999;  99US-0142920.        PR  28-SEP-1999;  99US-0156458.
PR  12-JUL-1999;  99US-0142977.        PR  29-SEP-1999;  99US-0156596.
PR  13-JUL-1999;  99US-0143542.        PR  04-OCT-1999;  99US-0157117.
PR  14-JUL-1999;  99US-0143624.        PR  05-OCT-1999;  99US-0157753.
PR  15-JUL-1999;  99US-0144005.        PR  06-OCT-1999;  99US-0157865.
PR  16-JUL-1999;  99US-0144085.        PR  07-OCT-1999;  99US-0158029.
PR  16-JUL-1999;  99US-0144086.        PR  08-OCT-1999;  99US-0158232.
PR  19-JUL-1999;  99US-0144325.        PR  12-OCT-1999;  99US-0158369.
PR  19-JUL-1999;  99US-0144331.        PR  13-OCT-1999;  99US-0159293.
PR  19-JUL-1999;  99US-0144332.        PR  13-OCT-1999;  99US-0159294.
PR  19-JUL-1999;  99US-0144333.        PR  13-OCT-1999;  99US-0159295.
PR  19-JUL-1999;  99US-0144334.        PR  14-OCT-1999;  99US-0159329.
PR  19-JUL-1999;  99US-0144335.        PR  14-OCT-1999;  99US-0159330.
PR  20-JUL-1999;  99US-0144352.        PR  14-OCT-1999;  99US-0159331.
PR  20-JUL-1999;  99US-0144632.        PR  14-OCT-1999;  99US-0159637.
PR  20-JUL-1999;  99US-0144884.        PR  14-OCT-1999;  99US-0159638.
PR  21-JUL-1999;  99US-0144814.        PR  18-OCT-1999;  99US-0159584.
PR  21-JUL-1999;  99US-0145086.        PR  21-OCT-1999;  99US-0160741.
PR  21-JUL-1999;  99US-0145088.        PR  21-OCT-1999;  99US-0160767.
PR  22-JUL-1999;  99US-0145085.        PR  21-OCT-1999;  99US-0160768.
PR  22-JUL-1999;  99US-0145087.        PR  21-OCT-1999;  99US-0160770.
PR  22-JUL-1999;  99US-0145089.        PR  22-OCT-1999;  99US-0160814.
PR  22-JUL-1999;  99US-0145192.        PR  22-OCT-1999;  99US-0160815.
PR  23-JUL-1999;  99US-0145145.        PR  22-OCT-1999;  99US-0160980.
PR  23-JUL-1999;  99US-0145218.        PR  22-OCT-1999;  99US-0160981.
PR  23-JUL-1999;  99US-0145224.        PR  25-OCT-1999;  99US-0160989.
PR  26-JUL-1999;  99US-0145276.        PR  25-OCT-1999;  99US-0161404.
PR  27-JUL-1999;  99US-0145913.        PR  25-OCT-1999;  99US-0161405.
PR  27-JUL-1999;  99US-0145918.        PR  26-OCT-1999;  99US-0161406.
PR  27-JUL-1999;  99US-0145919.        PR  26-OCT-1999;  99US-0161359.
PR  28-JUL-1999;  99US-0145951.        PR  26-OCT-1999;  99US-0161360.
PR  02-AUG-1999;  99US-0146388.        PR  28-OCT-1999;  99US-0161361.
PR  02-AUG-1999;  99US-0146389.        PR  28-OCT-1999;  99US-0161920.
PR  03-AUG-1999;  99US-0146386.        PR  28-OCT-1999;  99US-0161992.
PR  04-AUG-1999;  99US-0147204.        PR  29-OCT-1999;  99US-0161993.
PR  04-AUG-1999;  99US-0147302.        PR                99US-0162142.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
```

```
Query Match        15.2%;    Score 7;  DB 21;  Length 257;
Best Local Similarity 100.0%;   Pred. No. 41;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     25 GSGSLST 31
          |||||||
Db     88 GSGSLST 94

RESULT 7
```

```
AAG19902
ID   AAG19902 standard; Protein; 287 AA.
XX
AC   AAG19902;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 21880.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.        PR   18-JUN-1999; 99US-0139460.
PR   05-MAR-1999; 99US-0123180.        PR   18-JUN-1999; 99US-0139461.
PR   09-MAR-1999; 99US-0123548.        PR   18-JUN-1999; 99US-0139462.
PR   23-MAR-1999; 99US-0125788.        PR   18-JUN-1999; 99US-0139463.
PR   25-MAR-1999; 99US-0126264.        PR   18-JUN-1999; 99US-0139750.
PR   29-MAR-1999; 99US-0126785.        PR   18-JUN-1999; 99US-0139763.
PR   01-APR-1999; 99US-0127462.        PR   21-JUN-1999; 99US-0139817.
PR   06-APR-1999; 99US-0128234.        PR   22-JUN-1999; 99US-0139899.
PR   08-APR-1999; 99US-0128714.        PR   23-JUN-1999; 99US-0140353.
PR   16-APR-1999; 99US-0129845.        PR   23-JUN-1999; 99US-0140354.
PR   19-APR-1999; 99US-0130077.        PR   24-JUN-1999; 99US-0140695.
PR   21-APR-1999; 99US-0130449.        PR   28-JUN-1999; 99US-0140823.
PR   23-APR-1999; 99US-0130510.        PR   29-JUN-1999; 99US-0140991.
PR   23-APR-1999; 99US-0130891.        PR   30-JUN-1999; 99US-0141287.
PR   28-APR-1999; 99US-0131449.        PR   01-JUL-1999; 99US-0141842.
PR   30-APR-1999; 99US-0132048.        PR   01-JUL-1999; 99US-0142154.
PR   30-APR-1999; 99US-0132407.        PR   02-JUL-1999; 99US-0142055.
PR   04-MAY-1999; 99US-0132484.        PR   06-JUL-1999; 99US-0142390.
PR   05-MAY-1999; 99US-0132485.        PR   08-JUL-1999; 99US-0142803.
PR   06-MAY-1999; 99US-0132486.        PR   09-JUL-1999; 99US-0142920.
PR   06-MAY-1999; 99US-0132487.        PR   12-JUL-1999; 99US-0142977.
PR   07-MAY-1999; 99US-0132863.        PR   13-JUL-1999; 99US-0143542.
PR   11-MAY-1999; 99US-0134256.        PR   14-JUL-1999; 99US-0143624.
PR   14-MAY-1999; 99US-0134218.        PR   15-JUL-1999; 99US-0144005.
PR   14-MAY-1999; 99US-0134219.        PR   16-JUL-1999; 99US-0144085.
PR   14-MAY-1999; 99US-0134221.        PR   16-JUL-1999; 99US-0144086.
PR   14-MAY-1999; 99US-0134370.        PR   19-JUL-1999; 99US-0144325.
PR   18-MAY-1999; 99US-0134768.        PR   19-JUL-1999; 99US-0144331.
PR   19-MAY-1999; 99US-0134941.        PR   19-JUL-1999; 99US-0144332.
PR   20-MAY-1999; 99US-0135124.        PR   19-JUL-1999; 99US-0144333.
PR   21-MAY-1999; 99US-0135353.        PR   19-JUL-1999; 99US-0144334.
PR   24-MAY-1999; 99US-0135629.        PR   19-JUL-1999; 99US-0144335.
PR   25-MAY-1999; 99US-0136021.        PR   20-JUL-1999; 99US-0144352.
PR   27-MAY-1999; 99US-0136392.        PR   20-JUL-1999; 99US-0144632.
PR   28-MAY-1999; 99US-0136782.        PR   20-JUL-1999; 99US-0144884.
PR   01-JUN-1999; 99US-0137222.        PR   21-JUL-1999; 99US-0144814.
PR   03-JUN-1999; 99US-0137528.        PR   21-JUL-1999; 99US-0145086.
PR   04-JUN-1999; 99US-0137502.        PR   21-JUL-1999; 99US-0145088.
PR   07-JUN-1999; 99US-0137724.        PR   22-JUL-1999; 99US-0145085.
PR   08-JUN-1999; 99US-0138094.        PR   22-JUL-1999; 99US-0145087.
PR   10-JUN-1999; 99US-0138540.        PR   22-JUL-1999; 99US-0145089.
PR   10-JUN-1999; 99US-0138847.        PR   22-JUL-1999; 99US-0145192.
PR   14-JUN-1999; 99US-0139119.        PR   23-JUL-1999; 99US-0145145.
PR   16-JUN-1999; 99US-0139452.        PR   23-JUL-1999; 99US-0145218.
PR   16-JUN-1999; 99US-0139453.        PR   26-JUL-1999; 99US-0145224.
PR   17-JUN-1999; 99US-0139492.        PR   27-JUL-1999; 99US-0145276.
PR   18-JUN-1999; 99US-0139454.        PR   27-JUL-1999; 99US-0145913.
PR   18-JUN-1999; 99US-0139455.        PR   28-JUL-1999; 99US-0145918.
PR   18-JUN-1999; 99US-0139456.        PR   02-AUG-1999; 99US-0145919.
PR   18-JUN-1999; 99US-0139457.        PR   02-AUG-1999; 99US-0145951.
PR   18-JUN-1999; 99US-0139458.        PR   03-AUG-1999; 99US-0146386.
PR   18-JUN-1999; 99US-0139459.        PR   03-AUG-1999; 99US-0146388.
                                       PR   04-AUG-1999; 99US-0146389.
                                       PR   05-AUG-1999; 99US-0147038.
                                       PR   05-AUG-1999; 99US-0147204.
                                       PR   06-AUG-1999; 99US-0147302.
                                       PR   06-AUG-1999; 99US-0147192.
                                       PR   09-AUG-1999; 99US-0147260.
                                       PR   10-AUG-1999; 99US-0147303.
                                       PR   11-AUG-1999; 99US-0147416.
                                       PR   12-AUG-1999; 99US-0147493.
                                       PR   13-AUG-1999; 99US-0147935.
                                       PR   13-AUG-1999; 99US-0148171.
                                       PR   16-AUG-1999; 99US-0148319.
                                       PR   17-AUG-1999; 99US-0148341.
                                       PR   18-AUG-1999; 99US-0148555.
                                       PR   18-AUG-1999; 99US-0148684.
                                       PR   20-AUG-1999; 99US-0149368.
                                       PR   20-AUG-1999; 99US-0149175.
                                       PR   20-AUG-1999; 99US-0149426.
                                       PR   20-AUG-1999; 99US-0149722.
                                       PR   20-AUG-1999; 99US-0149723.
                                       PR   20-AUG-1999; 99US-0149929.
```

```
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158866.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match             15.2%;  Score 7;  DB 21;  Length 287;
Best Local Similarity   100.0%;  Pred. No. 46;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     25 GSGSLST 31
          |||||||
Db    118 GSGSLST 124

RESULT 8
AAR42086
ID   AAR42086 standard; protein; 558 AA.
XX
AC   AAR42086;
XX
DT   05-MAY-1994   (first entry)
XX
DE   Mouse RelB protein.
XX
KW   NF-KappaB/Rel/Dorsal; NRD; DNA binding region.
XX
OS   Mus musculus.
XX
PN   WO9320219-A.
XX
PD   14-OCT-1993.
XX
PF   01-APR-1993;   93WO-US03027.
XX
PR   06-APR-1992;   92US-0862987.
XX
PA   (USSH ) US SEC DEPT HEALTH.
XX
PI   Leonard WJ, Toledano MB;
XX
DR   WPI; 1993-336925/42.
XX
PT   Mutation of DNA binding region of NF-KB/rel/dorsal protein - to
PT   prevent or control binding to DNA for control of gene activation
XX
PS   Disclosure; Page 25-26; 52pp; English.
XX
CC   The sequence is that of the mouse RelB protein, the DNA binding
CC   region of which may be mutated by replacing one or more amino acids.
CC   This allows inhibition, prevention or control of DNA binding and
CC   thus activation of genes controlled by the protein can be controlled.
XX
SQ   Sequence   558 AA;

Query Match             15.2%;  Score 7;  DB 14;  Length 558;
Best Local Similarity   100.0%;  Pred. No. 85;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     16 TDELEII 22
          |||||||
Db     33 TDELEII 39

RESULT 9
AAG30638
ID   AAG30638 standard; Protein; 586 AA.
XX
AC   AAG30638;
XX
DT   17-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 36664.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126364.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
```

Wed Nov 13 14:19:03 2002

us-09-833-017b-2.oligo.rag

PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.

```
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.

Query Match             15.2%;    Score 7;  DB 21;  Length 586;
Best Local Similarity  100.0%;    Pred. No. 89;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    21 IIIGGSG 27
         |||||||
Db   423 IIIGGSG 429

RESULT 10
AAG30637
ID   AAG30637 standard; Protein; 612 AA.
XX
AC   AAG30637;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 36663.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   30-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   21-JUN-1999;  99US-0139750.
PR   22-JUN-1999;  99US-0139763.
PR   23-JUN-1999;  99US-0139817.
PR   23-JUN-1999;  99US-0139899.
PR   24-JUN-1999;  99US-0140353.
PR   28-JUN-1999;  99US-0140354.
PR   29-JUN-1999;  99US-0140695.
PR   30-JUN-1999;  99US-0140823.
PR   01-JUL-1999;  99US-0140991.
PR   02-JUL-1999;  99US-0141287.
PR   06-JUL-1999;  99US-0141842.
PR   08-JUL-1999;  99US-0142154.
PR   09-JUL-1999;  99US-0142055.
PR   12-JUL-1999;  99US-0142390.
PR   13-JUL-1999;  99US-0142803.
PR   14-JUL-1999;  99US-0142920.
PR   15-JUL-1999;  99US-0142977.
PR   16-JUL-1999;  99US-0143542.
PR   16-JUL-1999;  99US-0143624.
PR   19-JUL-1999;  99US-0144005.
PR   19-JUL-1999;  99US-0144085.
PR   19-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
PR   20-JUL-1999;  99US-0144332.
PR   20-JUL-1999;  99US-0144333.
PR   21-JUL-1999;  99US-0144334.
PR   21-JUL-1999;  99US-0144335.
PR                  99US-0144352.
PR                  99US-0144632.
PR                  99US-0144884.
PR                  99US-0144814.
PR                  99US-0145086.
PR                  99US-0145088.
```

```
Query Match      15.2%;    Score 7;   DB 21;   Length 612;
Best Local Similarity  100.0%;
Matches   7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    21 IIIGGSG 27
         |||||||
Db   449 IIIGGSG 455

RESULT 11
AAG30636
ID   AAG30636 standard; Protein; 618 AA.
XX
AC   AAG30636;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 36662.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
```

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 18-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 28-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 29-OCT-1999; 99US-0161992.
            99US-0161993.
            99US-0162142.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0127462.
PR 01-APR-1999; 99US-0127785.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
            99US-0134218.
```

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143524.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144333.
PR 20-JUL-1999; 99US-0144334.
PR 21-JUL-1999; 99US-0144335.
PR 21-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144632.
PR 22-JUL-1999; 99US-0144884.
PR 22-JUL-1999; 99US-0144814.
PR 22-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 23-JUL-1999; 99US-0145085.
PR 23-JUL-1999; 99US-0145087.
PR 23-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   15.2%;   Score 7; DB 21; Length 618;
Best Local Similarity 100.0%; Pred. No. 94;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   21 IIIGGSG 27
        |||||||
Db 455 IIIGGSG 461

RESULT 12
ABG17981
ID ABG17981 standard; Protein; 658 AA.
XX
AC ABG17981;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17972.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS82168.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 48340; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note:The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 658 AA;

Query Match   15.2%;   Score 7; DB 22; Length 658;
Best Local Similarity 100.0%; Pred. No. 99;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   17 DELEIII 23
        |||||||
Db 378 DELEIII 384

RESULT 13
ABB58240
ID ABB58240 standard; Protein; 767 AA.
XX
AC ABB58240;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1512.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02343.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 and cell-cell
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 1512; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL6176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 767 AA;

Query Match   15.2%;   Score 7; DB 22; Length 767;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      24 GGGGSLS 30
           |||||||
Db     304 GGGGSLS 310

RESULT 14
ABB70857
ID  ABB70857 standard; Protein; 929 AA.
XX
AC  ABB70857;
XX
DT  26-MAR-2002 (first entry)
XX
DE  Drosophila melanogaster polypeptide SEQ ID NO 39363.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical.
XX
OS  Drosophila melanogaster.
XX
PN  WO200171042-A2.
XX
PD  27-SEP-2001.
XX
PF  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
PR  11-JUL-2000; 2000US-0614150.
XX
PA  (PEKE ) PE CORP NY.
XX
PI  Venter JC, Adams M, Li PWD, Myers EW;
XX
DR  WPI; 2001-656860/75.
DR  N-PSDB; ABL14960.
XX
PT  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
PS  Disclosure; SEQ ID NO 39363; 21pp + Sequence Listing; English.
XX
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC  sequences (ABL01840-ABL16175) and the encoded proteins
CC  (ABB57737-ABB72072).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence  929 AA;

Query Match      15.2%;  Score 7;  DB 22;  Length 929;
Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
Matches    7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      39 SFTQALG 45
           |||||||
Db     103 SFTQALG 109

RESULT 15
ABB63880
ID  ABB63880 standard; Protein; 2771 AA.
XX
AC  ABB63880;
XX
DT  26-MAR-2002 (first entry)
XX
DE  Drosophila melanogaster polypeptide SEQ ID NO 18432.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical.
XX
OS  Drosophila melanogaster.
XX
PN  WO200171042-A2.
XX
PD  27-SEP-2001.
XX
PF  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
PR  11-JUL-2000; 2000US-0614150.
XX
PA  (PEKE ) PE CORP NY.
XX
PI  Venter JC, Adams M, Li PWD, Myers EW;
XX
DR  WPI; 2001-656860/75.
DR  N-PSDB; ABL07983.
XX
PT  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
PS  Disclosure; SEQ ID NO 18432; 21pp + Sequence Listing; English.
XX
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC  sequences (ABL01840-ABL16175) and the encoded proteins
CC  (ABB57737-ABB72072).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence  2771 AA;

Query Match      15.2%;  Score 7;  DB 22;  Length 2771;
Best Local Similarity  100.0%;  Pred. No. 3.8e+02;
Matches    7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      24 GGGGSLS 30
           |||||||
Db    1898 GGGGSLS 1904

Search completed: November  5, 2002, 11:04:58
Job time : 78.3881 secs
```

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2002, 11:03:27 ; Search time 15.1045 Seconds
                              (without alignments)
                              74.387 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 46
Sequence: 1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: OLIGO
                Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size :   0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   Issued_Patents_AA:*
                1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15.2	7	558	5	PCT-US93-03027-2	Sequence 2, Appli
2	15.2	7	767	4	US-08-836-567-8	Sequence 8, Appli
3	13.0	6	17	1	US-08-370-567-18	Sequence 18, Appli
4	13.0	6	17	1	US-08-438-759-18	Sequence 18, Appli
5	13.0	6	17	5	PCT-US94-05684-18	Sequence 18, Appl
6	13.0	6	114	2	US-08-652-558-52	Sequence 52, Appli
7	13.0	6	134	4	US-09-367-953A-120	Sequence 120, App
8	13.0	6	134	4	US-09-367-953B-121	Sequence 121, App
9	13.0	6	282	4	US-08-482-918-56	Sequence 56, Appli
10	13.0	6	282	4	US-09-224-681-56	Sequence 56, Appl
11	13.0	6	282	4	US-08-336-728A-56	Sequence 56, Appl
12	13.0	6	322	4	US-09-383-586-33	Sequence 33, Appl
13	13.0	6	397	3	US-09-079-415-2	Sequence 2, Appli
14	13.0	6	398	1	US-08-328-314-2	Sequence 2, Appli
15	13.0	6	398	1	US-08-731-045-2	Sequence 2, Appli
16	13.0	6	412	4	US-08-208-007A-12	Sequence 12, Appli
17	13.0	6	412	4	US-08-974-691-4	Sequence 4, Appli
18	13.0	6	439	4	US-09-383-586-32	Sequence 32, Appl
19	13.0	6	446	4	US-09-268-992-51	Sequence 51, Appl
20	13.0	6	466	4	US-09-268-992-64	Sequence 64, Appl
21	13.0	6	477	4	US-09-446-504-19	Sequence 19, Appli
22	13.0	6	477	4	US-09-712-266-19	Sequence 19, Appl
23	13.0	6	477	4	US-09-268-992-4	Sequence 4, Appli
24	13.0	6	493	4	US-09-177-349-5	Sequence 5, Appli
25	13.0	6	495	4	US-09-268-992-2	Sequence 2, Appli
26	13.0	6	499	2	US-07-952-853-6	Sequence 6, Appli
27	13.0	6	499	2	US-08-914-848-6	Sequence 6, Appli
28		6	13.0	508	2	US-08-818-024-3	Sequence 3, Appli
29		6	13.0	508	4	US-09-334-775A-3	Sequence 3, Appli
30		6	13.0	529	4	US-09-383-586-31	Sequence 31, Appli
31		6	13.0	534	2	US-08-691-814B-8	Sequence 8, Appli
32		6	13.0	538	2	US-08-541-759B-2	Sequence 2, Appli
33		6	13.0	547	2	US-08-467-822-35	Sequence 35, Appl
34		6	13.0	547	4	US-08-432-697-35	Sequence 35, Appl
35		6	13.0	547	2	US-08-466-248-35	Sequence 35, Appl
36		6	13.0	573	2	US-08-706-209-1	Sequence 1, Appli
37		6	13.0	573	3	US-08-981-787-1	Sequence 1, Appli
38		6	13.0	573	4	US-08-461-722-1	Sequence 1, Appli
39		6	13.0	573	4	US-08-336-251-1	Sequence 1, Appli
40		6	13.0	573	5	PCT-US94-06362-1	Sequence 1, Appli
41		6	13.0	573	5	PCT-US96-11373-1	Sequence 1, Appli
42		6	13.0	573	5	PCT-US96-11375-1	Sequence 1, Appl
43		6	13.0	635	2	US-09-014-969-11	Sequence 11, Appl
44		6	13.0	719	4	US-09-386-607-2	Sequence 2, Appli
45		6	13.0	887	1	US-07-596-467-2	Sequence 2, Appli
46		6	13.0	887	1	US-07-934-374-2	Sequence 2, Appli
47		6	13.0	887	1	US-07-788-861C-4	Sequence 4, Appli
48		6	13.0	888	4	US-09-268-140-4	Sequence 4, Appli
49		6	13.0	893	4	US-07-977-934-4	Sequence 4, Appli
50		6	13.0	893	4	US-09-458-819-4	Sequence 10, Appl
51		6	13.0	893	4	US-09-105-697-10	Sequence 4, Appli
52		6	13.0	893	5	PCT-US91-07035-4	Sequence 17, Appl
53		6	13.0	913	3	US-08-911-853-17	Sequence 17, Appli
54		6	13.0	913	4	US-09-479-409-17	Sequence 17, Appl
55		6	13.0	913	4	US-09-479-453-17	Sequence 7, Appli
56		6	13.0	1968	1	US-07-745-206A-7	Sequence 45, Appl
57		6	13.0	1968	2	US-08-455-543A-45	Sequence 45, Appl
58		6	13.0	1968	2	US-08-223-305C-45	Sequence 7, Appl
59		6	13.0	1968	2	US-08-311-363-7	Sequence 2, Appli
60		6	13.0	2161	1	US-07-745-206A-2	Sequence 49, Appli
61		6	13.0	2161	1	US-08-455-543A-49	Sequence 51, Appli
62		6	13.0	2161	2	US-08-455-543A-51	Sequence 49, Appli
63		6	13.0	2161	2	US-08-223-305C-49	Sequence 51, Appli
64		6	13.0	2161	2	US-08-223-305C-51	Sequence 2, Appli
65		6	13.0	2161	2	US-08-311-363-2	Sequence 11, Appli
66		5	10.9	8	1	US-08-148-910-11	Sequence 11, Appli
67		5	10.9	8	1	US-08-448-937A-11	Sequence 4, Appli
68		5	10.9	8	4	US-09-139-802-4	Sequence 11, Appli
69		5	10.9	14	4	US-08-849-488-16	Sequence 16, Appli
70		5	10.9	15	1	US-07-664-989B-11	Sequence 11, Appli
71		5	10.9	15	2	US-07-664-989B-18	Sequence 18, Appli
72		5	10.9	15	2	US-08-726-464B-30	Sequence 30, Appli
73		5	10.9	15	2	US-08-933-983-74	Sequence 74, Appli
74		5	10.9	16	2	US-08-480-190-261	Sequence 261, App
75		5	10.9	16	2	US-08-474-696A-7	Sequence 7, Appli
76		5	10.9	16	2	US-08-474-696A-8	Sequence 8, Appli
77		5	10.9	16	2	US-08-488-379-261	Sequence 261, App
78		5	10.9	16	5	PCT-US93-07545-261	Sequence 261, App
79		5	10.9	19	2	US-08-474-696A-1	Sequence 1, Appli
80		5	10.9	19	2	US-08-474-696A-3	Sequence 3, Appli
81		5	10.9	23	1	US-08-383-753-82	Sequence 82, Appli
82		5	10.9	23	2	US-08-586-772-82	Sequence 82, Appl
83		5	10.9	23	3	US-08-959-512-82	Sequence 82, Appl
84		5	10.9	23	4	US-09-512-983-82	Sequence 82, Appl
85		5	10.9	26	1	US-08-524-677-13	Sequence 13, Appl
86		5	10.9	29	4	US-09-227-357-371	Sequence 371, App
87		5	10.9	31	1	US-08-190-130-5	Sequence 5, Appli
88		5	10.9	31	4	US-09-177-249-21	Sequence 21, Appl
89		5	10.9	34	3	US-08-491-954-36	Sequence 36, Appli
90		5	10.9	35	3	US-08-491-954-39	Sequence 39, Appli
91		5	10.9	38	1	US-08-987-272A-18	Sequence 18, Appl
92		5	10.9	38	3	US-08-491-954-37	Sequence 37, Appli
93		5	10.9	41	3	US-08-491-954-38	Sequence 38, Appli
94		5	10.9	45	1	US-08-462-128-39	Sequence 39, Appl
95		5	10.9	45	3	US-08-463-180-39	Sequence 39, Appl
96		5	10.9	49	3	US-08-469-318-194	Sequence 194, App
97		5	10.9	49	3	US-08-468-609A-194	Sequence 194, App
98		5	10.9	49	5	PCT-US95-01185-194	Sequence 194, App
99		5	10.9	51	4	US-08-818-112-104	Sequence 104, App
100		5	10.9	51	4	US-08-818-111-99	Sequence 99, Appl

#			N	Sequence ID	Label
101	10.9	5	4	US-09-056-556-104	Sequence 104, App
102	10.9	5	1	US-08-346-849-8	Sequence 8, Appli
103	10.9	5	2	US-08-293-284A-8	Sequence 8, Appli
104	10.9	5	1	US-08-485-455D-35	Sequence 35, Appl
105	10.9	5	2	US-08-482-130C-35	Sequence 35, Appl
106	10.9	5	2	US-08-484-211C-35	Sequence 35, Appl
107	10.9	5	3	US-08-906-769-35	Sequence 35, Appl
108	10.9	5	3	US-08-906-616-35	Sequence 35, Appl
109	10.9	5	4	US-08-817-795-35	Sequence 35, Appl
110	10.9	5	4	US-08-485-443B-35	Sequence 35, Appl
111	10.9	5	4	US-08-639-075A-35	Sequence 35, Appl
112	10.9	5	4	US-09-012-431-35	Sequence 35, Appl
113	10.9	5	4	US-09-012-692-35	Sequence 35, Appl
114	10.9	5	4	US-08-906-613-35	Sequence 35, Appl
115	10.9	5	5	PCT-US95-14442A-35	Sequence 35, Appl
116	10.9	5	3	US-08-469-318-195	Sequence 195, App
117	10.9	5	3	US-08-468-609A-195	Sequence 195, App
118	10.9	5	2	PCT-US95-01185-195	Sequence 195, App
119	10.9	5	2	US-08-245-511-36	Sequence 36, Appl
120	10.9	5	2	US-08-600-993A-36	Sequence 36, Appl
121	10.9	5	2	US-09-042-012-8	Sequence 8, Appli
122	10.9	5	4	US-09-305-086-2	Sequence 2, Appl
123	10.9	5	4	US-09-457-324-8	Sequence 8, Appli
124	10.9	5	1	US-07-987-272A-17	Sequence 17, Appl
125	10.9	5	3	US-08-680-726A-60	Sequence 60, Appl
126	10.9	5	3	US-09-092-409-60	Sequence 60, Appl
127	10.9	5	2	US-08-691-814B-29	Sequence 29, Appl
128	10.9	5	2	US-08-562-114B-24	Sequence 24, Appl
129	10.9	5	3	US-08-729-594A-24	Sequence 24, Appl
130	10.9	5	1	US-07-987-272A-1	Sequence 1, Appl
131	10.9	5	3	US-08-906-769-95	Sequence 95, Appl
132	10.9	5	3	US-08-906-616-95	Sequence 95, Appl
133	10.9	5	4	US-08-817-795-95	Sequence 95, Appl
134	10.9	5	4	US-08-639-075A-95	Sequence 95, Appl
135	10.9	5	4	US-09-012-431-95	Sequence 95, Appl
136	10.9	5	4	US-09-012-692-95	Sequence 95, Appl
137	10.9	5	4	US-08-906-613-95	Sequence 95, Appl
138	10.9	5	4	US-08-936-165A-95	Sequence 95, Appl
139	10.9	5	5	PCT-US95-14442A-95	Sequence 287, App
140	10.9	5	1	US-07-987-272A-14	Sequence 14, Appl
141	10.9	5	3	US-08-816-977-4	Sequence 4, Appl
142	10.9	5	2	US-08-465-640-2	Sequence 2, Appl
144	10.9	5	3	US-08-300-928C-10	Sequence 10, Appl
145	10.9	5	3	US-08-430-944D-10	Sequence 10, Appl
146	10.9	5	3	US-08-430-014-10	Sequence 10, Appl
147	10.9	5	3	US-08-816-977-23	Sequence 23, Appl
148	10.9	5	3	US-08-431-184-10	Sequence 10, Appl
149	10.9	5	1	US-08-858-207A-431	Sequence 431, App
150	10.9	5	3	US-08-369-796-16	Sequence 16, Appl
151	10.9	5	3	US-08-852-091-16	Sequence 16, Appl
152	10.9	5	5	PCT-US95-17025-16	Sequence 16, Appl
153	10.9	5	1	US-07-662-193-5	Sequence 5, Appl
154	10.9	5	3	US-08-807-529A-6	Sequence 6, Appli
155	10.9	5	3	US-08-300-928C-8	Sequence 8, Appli
156	10.9	5	3	US-08-430-944D-8	Sequence 8, Appli
157	10.9	5	3	US-08-430-014-8	Sequence 8, Appli
158	10.9	5	3	US-08-431-184-8	Sequence 8, Appli
159	10.9	5	5	PCT-US93-02462-6	Sequence 4, Appli
160	10.9	5	3	US-08-300-928C-6	Sequence 6, Appli
161	10.9	5	3	US-08-430-944D-6	Sequence 6, Appli
162	10.9	5	3	US-08-430-014-6	Sequence 6, Appli
163	10.9	5	3	US-08-431-184-6	Sequence 6, Appli
164	10.9	5	1	US-08-139-862-4	Sequence 4, Appli
165	10.9	5	2	US-09-046-479-2	Sequence 2, Appli
166	10.9	5	2	US-08-446-945-27	Sequence 27, Appl
167	10.9	5	2	US-08-082-623-3	Sequence 3, Appli
168	10.9	5	4	US-08-983-607-47	Sequence 47, Appl
169	10.9	5	4	US-08-789-333F-41	Sequence 41, Appl
170	10.9	5	3	US-08-751-359-13	Sequence 13, Appl
171	10.9	5	3	US-08-907-146-13	Sequence 13, Appl
172	10.9	5	3	US-08-751-359-12	Sequence 12, Appl
173	10.9	5	4	US-08-907-146-12	Sequence 12, Appl

#			N	Sequence ID	Label
174	10.9	5	1	US-08-253-155A-35	Sequence 35, Appl
175	10.9	5	3	US-08-789-333F-43	Sequence 43, Appl
176	10.9	5	3	US-09-014-574-1	Sequence 1, Appli
177	10.9	5	1	US-09-553-498-4	Sequence 4, Appli
178	10.9	5	2	US-08-387-942C-49	Sequence 49, Appl
179	10.9	5	4	US-09-247-155-110	Sequence 110, App
180	10.9	5	4	US-08-074-121-5	Sequence 5, Appl
181	10.9	5	4	US-09-181-183-2	Sequence 2, Appl
182	10.9	5	4	US-08-858-207A-334	Sequence 334, App
183	10.9	5	4	US-09-277-700-2	Sequence 2, Appli
184	10.9	5	5	PCT-US94-06447-5	Sequence 5, Appli
185	10.9	5	4	US-09-025-769B-372	Sequence 372, App
186	10.9	5	5	US-09-025-769B-373	Sequence 373, App
187	10.9	5	3	US-08-772-270A-7	Sequence 7, Appl
188	10.9	5	3	US-09-027-449-53	Sequence 53, Appl
189	10.9	5	3	US-08-804-444A-53	Sequence 53, Appl
190	10.9	5	4	US-09-026-985-53	Sequence 53, Appl
191	10.9	5	4	US-09-025-769B-282	Sequence 282, App
192	10.9	5	4	US-09-087-465-30	Sequence 30, Appl
193	10.9	5	4	US-08-720-625-4	Sequence 4, Appli
194	10.9	5	2	US-08-764-563-5	Sequence 39, Appl
195	10.9	5	3	US-08-557-309B-39	Sequence 39, Appl
196	10.9	5	3	US-08-834-306-39	Sequence 39, Appl
197	10.9	5	4	US-08-993-674A-39	Sequence 42, Appl
198	10.9	5	4	US-08-789-333F-42	Sequence 42, Appl
199	10.9	5	4	US-08-789-333F-44	Sequence 44, Appl
200	10.9	5	1	US-08-044-621D-32	Sequence 32, Appl
201	10.9	5	1	US-08-709-912-15	Sequence 15, Appl
202	10.9	5	3	US-09-047-370-15	Sequence 15, Appl
203	10.9	5	4	US-08-764-563-4	Sequence 4, Appli
204	10.9	5	4	US-08-862-12-2	Sequence 8, Appli
205	10.9	5	4	US-08-328-254-8	Sequence 35, Appl
206	10.9	5	4	US-08-230-637-35	Sequence 35, Appl
207	10.9	5	3	US-08-764-563-1	Sequence 1, Appli
208	10.9	5	4	US-08-720-625-2	Sequence 2, Appli
209	10.9	5	2	US-08-685-992-1	Sequence 1, Appli
210	10.9	5	2	US-09-144-925-1	Sequence 1, Appli
211	10.9	5	2	US-08-484-126-7	Sequence 7, Appl
212	10.9	5	1	US-08-602-262-2	Sequence 2, Appli
213	10.9	5	3	US-09-004-716-2	Sequence 2, Appli
214	10.9	5	6	5189147-8	Patent No. 5189147-8
215	10.9	5	2	US-09-342-084-10	Sequence 10, Appl
216	10.9	5	2	US-09-808-550-32	Sequence 32, Appl
217	10.9	5	1	US-08-276-852-34	Sequence 34, Appl
218	10.9	5	1	US-08-133-011-16	Sequence 16, Appl
219	10.9	5	1	US-08-322-730A-16	Sequence 16, Appl
220	10.9	5	1	US-08-387-874-16	Sequence 16, Appl
221	10.9	5	1	US-08-899-575-34	Sequence 34, Appl
222	10.9	5	1	US-08-899-575-34	Sequence 34, Appl
223	10.9	5	2	US-08-383-619-16	Sequence 16, Appl
224	10.9	5	1	US-08-907-739-16	Sequence 16, Appl
225	10.9	5	5	PCT-US93-08364-16	Sequence 16, Appl
226	10.9	5	5	PCT-US95-08743-34	Sequence 34, Appl
227	10.9	5	3	US-08-737-129A-2	Sequence 2, Appli
228	10.9	5	3	US-08-486-099-110	Sequence 110, App
229	10.9	5	3	US-08-360-107A-120	Sequence 120, App
230	10.9	5	3	US-08-484-223B-110	Sequence 110, App
231	10.9	5	3	US-08-919-597-110	Sequence 110, App
232	10.9	5	3	US-08-475-668A-110	Sequence 110, App
233	10.9	5	3	US-08-485-551A-110	Sequence 110, App
234	10.9	5	3	US-08-471-913A-110	Sequence 110, App
235	10.9	5	4	US-08-485-264A-110	Sequence 110, App
236	10.9	5	4	US-08-474-349A-110	Sequence 110, App
237	10.9	5	2	US-08-277-231A-14	Sequence 14, Appl
238	10.9	5	2	US-08-473-750-3	Sequence 3, Appli
239	10.9	5	2	US-08-477-326-3	Sequence 3, Appli
240	10.9	5	2	US-08-277-231A-13	Sequence 13, Appl
241	10.9	5	2	US-08-477-750-2	Sequence 2, Appli
242	10.9	5	2	US-08-477-326-2	Sequence 2, Appli
243	10.9	5	2	US-08-277-231A-2	Sequence 2, Appli
244	10.9	5	2	US-08-473-750-5	Sequence 5, Appli
245	10.9	5	2	US-08-477-326-5	Sequence 5, Appli
246	10.9	5	5	PCT-US94-14106-59	Sequence 59, Appl

No.					Reference	Sequence
247	10.9	5	218	4	US-09-227-357-368	Sequence 368, App
248	10.9	5	219	5	US-08-152-019A-31	Sequence 31, Appl
249	10.9	5	219	1	US-08-152-019A-32	Sequence 32, Appl
250	10.9	5	219	1	US-08-266-451B-24	Sequence 24, Appl
251	10.9	5	219	2	US-08-460-309-18	Sequence 18, Appl
252	10.9	5	219	2	US-08-748-725-24	Sequence 24, Appl
253	10.9	5	219	2	US-08-125-077-18	Sequence 18, Appl
254	10.9	5	220	4	US-09-052-089A-3	Sequence 3, Appl
255	10.9	5	222	2	US-08-737-129A-6	Sequence 6, Appl
256	10.9	5	222	4	US-09-181-183-28	Sequence 28, Appl
257	10.9	5	222	4	US-09-277-700-28	Sequence 28, Appl
258	10.9	5	223	4	US-09-129-030-30	Sequence 30, Appl
259	10.9	5	223	5	PCT-US94-14106-51	Sequence 51, Appl
260	10.9	5	223	5	PCT-US94-14106-55	Sequence 55, Appl
261	10.9	5	227	4	US-08-213-419B-13	Sequence 13, Appl
262	10.9	5	228	4	US-09-181-958-2	Sequence 2, Appl
263	10.9	5	229	2	US-08-121-436A-4	Sequence 4, Appli
264	10.9	5	230	4	US-08-808-550-34	Sequence 34, Appl
265	10.9	5	232	2	US-08-829-876-103	Sequence 103, Appl
266	10.9	5	233	2	US-08-829-876-101	Sequence 101, App
267	10.9	5	233	2	US-08-829-876-105	Sequence 105, App
268	10.9	5	233	4	US-08-438-745-15	Sequence 15, Appl
269	10.9	5	233	4	US-08-438-745-17	Sequence 17, Appl
270	10.9	5	233	4	US-09-219-019-15	Sequence 15, Appl
271	10.9	5	233	4	US-09-219-019-17	Sequence 17, Appl
272	10.9	5	233	4	PCT-US94-05669A-15	Sequence 15, Appl
273	10.9	5	233	5	PCT-US94-05669A-17	Sequence 17, Appl
274	10.9	5	234	4	US-09-040-483-5	Sequence 5, Appli
275	10.9	5	235	4	US-08-438-745-13	Sequence 13, Appl
276	10.9	5	235	4	US-09-219-019-13	Sequence 13, Appl
277	10.9	5	235	5	PCT-US94-05669A-13	Sequence 13, Appl
278	10.9	5	236	3	US-08-158-682A-4	Sequence 4, Appl
279	10.9	5	236	1	US-08-687-895-4	Sequence 4, Appli
280	10.9	5	236	1	US-08-816-241-4	Sequence 4, Appli
281	10.9	5	236	2	US-09-040-482-4	Sequence 4, Appli
282	10.9	5	236	3	US-09-128-395-4	Sequence 4, Appli
283	10.9	5	239	2	US-08-916-902A-1	Sequence 1, Appli
284	10.9	5	239	2	US-09-213-389-1	Sequence 1, Appli
285	10.9	5	240	4	US-09-194-905-8	Sequence 8, Appli
286	10.9	5	243	3	US-09-286-690-10	Sequence 10, Appl
287	10.9	5	246	2	US-08-553-497A-24	Sequence 24, Appl
288	10.9	5	247	3	US-09-100-804-13	Sequence 13, Appl
289	10.9	5	247	5	PCT-US94-07659-2	Sequence 2, Appli
290	10.9	5	248	4	US-08-944-483-71	Sequence 71, Appl
291	10.9	5	251	4	US-08-933-750C-44	Sequence 44, Appl
292	10.9	5	251	2	US-09-234-613-44	Sequence 44, Appl
293	10.9	5	251	4	US-08-944-483-28	Sequence 28, Appl
294	10.9	5	254	2	US-08-792-824-4	Sequence 4, Appli
295	10.9	5	254	2	US-08-792-824-7	Sequence 7, Appli
296	10.9	5	254	2	US-08-792-824-10	Sequence 10, Appl
297	10.9	5	254	2	US-08-792-824-13	Sequence 13, Appl
298	10.9	5	261	4	US-08-927-219-55	Sequence 55, Appl
299	10.9	5	264	2	US-08-904-284-3	Sequence 3, Appli
300	10.9	5	268	1	US-08-431-387-4	Sequence 4, Appli
301	10.9	5	268	2	US-07-857-224B-33	Sequence 33, Appl
302	10.9	5	269	1	US-08-203-716-3	Sequence 3, Appli
303	10.9	5	269	2	US-08-440-179-3	Sequence 3, Appli
304	10.9	5	269	2	US-08-432-693-11	Sequence 11, Appli
305	10.9	5	269	4	US-08-599-895-11	Sequence 11, Appl
306	10.9	5	269	2	US-08-483-806-2	Sequence 2, Appli
307	10.9	5	269	3	US-09-211-290-11	Sequence 11, Appl
308	10.9	5	269	3	US-09-030-613-11	Sequence 11, Appl
309	10.9	5	269	3	US-09-322-676-11	Sequence 11, Appl
310	10.9	5	269	4	US-09-039-657-3	Sequence 3, Appli
311	10.9	5	269	4	US-08-748-547-4	Sequence 4, Appli
312	10.9	5	269	4	US-09-055-095-4	Sequence 4, Appli
313	10.9	5	269	4	US-09-466-036A-11	Sequence 11, Appl
314	10.9	5	269	5	US-09-451-905-11	Sequence 11, Appl
315	10.9	5	269	1	PCT-US91-02339-1	Sequence 1, Appli
316	10.9	5	270	1	US-08-347-471-4	Sequence 4, Appli
317	10.9	5	270	2	US-09-809-494A-2	Sequence 2, Appli
318	10.9	5	270	4	US-09-352-302-2	Sequence 2, Appli
319	10.9	5	273	2	US-08-809-494A-4	Sequence 4, Appli
320	10.9	5	273	4	US-09-352-302-4	Sequence 4, Appli
321	10.9	5	274	3	US-08-491-954-4	Sequence 4, Appli
322	10.9	5	276	2	US-08-701-95-1	Sequence 1, Appli
323	10.9	5	276	3	US-09-134-591-1	Sequence 1, Appli
324	10.9	5	278	3	US-08-491-954-2	Sequence 2, Appli
325	10.9	5	280	1	US-08-434-255-8	Sequence 8, Appli
326	10.9	5	280	1	US-08-459-967-8	Sequence 8, Appli
327	10.9	5	280	1	US-08-460-327-8	Sequence 8, Appli
328	10.9	5	280	1	US-08-459-871-8	Sequence 8, Appli
329	10.9	5	280	4	US-09-024-532-2	Sequence 2, Appli
330	10.9	5	280	4	US-09-104-623A-2	Sequence 2, Appli
331	10.9	5	282	5	PCT-US94-09752-3	Sequence 3, Appli
332	10.9	5	284	1	US-08-457-245-3	Sequence 3, Appli
333	10.9	5	287	3	US-09-173-581-2	Sequence 2, Appli
334	10.9	5	287	4	US-08-862-124-17	Sequence 17, Appl
335	10.9	5	287	3	US-09-420-915-2	Sequence 2, Appli
336	10.9	5	288	3	US-09-100-804-14	Sequence 14, Appl
337	10.9	5	292	3	US-08-851-190-3	Sequence 3, Appli
338	10.9	5	293	4	US-08-438-745-4	Sequence 4, Appli
339	10.9	5	293	4	US-08-438-745-6	Sequence 6, Appli
340	10.9	5	293	4	US-09-219-019-4	Sequence 4, Appli
341	10.9	5	293	4	US-09-219-019-6	Sequence 6, Appli
342	10.9	5	293	5	PCT-US94-05669A-4	Sequence 4, Appli
343	10.9	5	293	5	PCT-US94-05669A-6	Sequence 6, Appli
344	10.9	5	296	2	US-08-836-854-4	Sequence 4, Appli
345	10.9	5	297	3	US-08-534-910B-7	Sequence 7, Appli
346	10.9	5	297	1	US-08-534-910B-10	Sequence 10, App
347	10.9	5	297	4	US-09-173-581-1	Sequence 1, Appli
348	10.9	5	297	4	US-09-420-915-1	Sequence 1, Appli
349	10.9	5	299	3	US-08-986-765-2	Sequence 2, Appli
350	10.9	5	299	4	US-08-935-307-2	Sequence 2, Appli
351	10.9	5	299	4	US-09-524-641-2	Sequence 2, Appli
352	10.9	5	299	4	US-08-965-130-2	Sequence 2, Appli
353	10.9	5	300	1	US-08-148-910-1	Sequence 1, Appli
354	10.9	5	300	4	US-08-448-937A-1	Sequence 1, Appli
355	10.9	5	300	4	US-08-158-735A-16	Sequence 16, App
356	10.9	5	300	3	US-08-158-735A-1	Sequence 17, App
357	10.9	5	301	3	US-08-469-318-142	Sequence 142, App
358	10.9	5	301	5	US-08-468-609A-142	Sequence 142, App
359	10.9	5	301	4	PCT-US95-01185-142	Sequence 142, App
360	10.9	5	304	4	US-08-862-124-14	Sequence 14, App
361	10.9	5	310	1	US-08-363-255-6	Sequence 6, Appli
362	10.9	5	312	1	US-08-247-908A-2	Sequence 2, Appli
363	10.9	5	312	2	US-08-453-942-2	Sequence 2, Appli
364	10.9	5	312	2	US-08-926-885A-2	Sequence 2, Appli
365	10.9	5	312	5	PCT-US94-05290-2	Sequence 2, Appli
366	10.9	5	320	2	US-08-579-940-8	Sequence 8, Appli
367	10.9	5	320	3	US-08-964-308-2	Sequence 2, Appli
368	10.9	5	320	3	US-08-964-313-2	Sequence 2, Appli
369	10.9	5	322	3	US-08-069-138-2	Sequence 2, Appli
370	10.9	5	322	3	US-08-964-308-8	Sequence 8, Appli
371	10.9	5	322	3	US-08-964-313-8	Sequence 8, Appli
372	10.9	5	324	2	US-09-069-138-8	Sequence 8, Appli
373	10.9	5	325	4	US-08-579-940-7	Sequence 7, Appli
374	10.9	5	328	2	US-08-878-474-9	Sequence 9, Appli
375	10.9	5	328	2	US-08-229-287-4	Sequence 4, Appli
376	10.9	5	331	1	US-08-977-767-1	Sequence 1, Appli
377	10.9	5	331	1	US-08-094-533B-10	Sequence 10, Appl
378	10.9	5	331	4	US-08-276-860A-10	Sequence 10, Appl
379	10.9	5	331	1	US-08-961-083-212	Sequence 212, App
380	10.9	5	331	1	US-09-452-370-10	Sequence 10, Appl
381	10.9	5	331	4	US-08-444-393-10	Sequence 10, Appl
382	10.9	5	331	2	US-08-799-913-10	Sequence 10, Appl
383	10.9	5	331	4	US-08-711-893-10	Sequence 10, Appl
384	10.9	5	331	3	US-09-150-201-10	Sequence 10, Appl
385	10.9	5	331	4	US-08-150-200-10	Sequence 10, Appl
386	10.9	5	331	4	PCT-US94-08119-10	Sequence 10, Appl
387	10.9	5	331	5	PCT-US94-08120-10	Sequence 10, Appl
388	10.9	5	331	5	PCT-US94-12913A-10	Sequence 10, Appl
389	10.9	5	331	5	PCT-US95-15353-39	Sequence 39, Appl
390	10.9	5	331	5	PCT-US95-15353-41	Sequence 41, Appl

No.	C			Patent / US	Sequence
393	3	10.9	5	US-08-469-318-143	Sequence 143, App
394	3	10.9	5	US-08-468-609A-143	Sequence 143, Appl
395	5	10.9	5	PCT-US95-01185-143	Sequence 143, App
396	3	10.9	5	US-08-469-318-148	Sequence 148, App
397	3	10.9	5	US-08-468-609A-148	Sequence 148, Appl
398	3	10.9	5	US-08-480-070C-9	Sequence 9, Appli
399	3	10.9	5	US-08-829-525-9	Sequence 9, Appli
400	4	10.9	5	US-08-609-583A-9	Sequence 9, Appli
401	3	10.9	5	US-08-937-399-9	Sequence 9, Appli
402	5	10.9	5	PCT-US95-01185-148	Sequence 148, App
403	3	10.9	5	US-08-978-456-2	Sequence 2, Appli
404	3	10.9	5	US-09-369-700-2	Sequence 2, Appli
405	1	10.9	5	US-08-463-090B-11	Sequence 11, Appl
406	4	10.9	5	US-08-961-083-192	Sequence 192, Appl
407	2	10.9	5	US-08-446-345-40	Sequence 40, Appl
408	3	10.9	5	US-08-469-318-139	Sequence 139, App
409	3	10.9	5	US-08-468-609A-139	Sequence 139, Appl
410	3	10.9	5	US-08-468-609A-151	Sequence 151, Appl
411	3	10.9	5	US-08-469-318-151	Sequence 151, Appl
412	5	10.9	5	PCT-US95-01185-139	Sequence 139, App
413	5	10.9	5	PCT-US95-01185-151	Sequence 151, App
414	1	10.9	5	US-08-073-807A-17	Sequence 17, Appl
415	1	10.9	5	US-08-597-236-11	Sequence 11, Appl
416	1	10.9	5	US-08-746-682A-11	Sequence 11, Appli
417	4	10.9	5	US-09-364-230-4	Sequence 4, Appli
418	1	10.9	5	US-08-537-434-1	Sequence 1, Appli
419	4	10.9	5	US-08-096-181A-8	Sequence 8, Appli
420	5	10.9	5	PCT-US94-0326-8	Sequence 2, Appli
421	4	10.9	5	US-08-445-090-2	Sequence 2, Appli
422	4	10.9	5	US-09-286-691-14	Sequence 14, Appl
423	4	10.9	5	US-09-687-147-14	Sequence 14, Appli
424	1	10.9	5	US-08-484-105-20	Sequence 20, Appl
425	4	10.9	5	US-08-484-106-20	Sequence 20, Appli
426	2	10.9	5	US-08-646-590B-36	Sequence 36, Appl
427	3	10.9	5	US-09-412-184-36	Sequence 36, Appli
428	4	10.9	5	US-08-213-419B-15	Sequence 21, Appl
429	1	10.9	5	US-08-700-359-21	Sequence 21, Appl
430	1	10.9	5	US-08-434-255-6	Sequence 6, Appli
431	1	10.9	5	US-08-459-967-6	Sequence 6, Appli
432	1	10.9	5	US-08-459-871-6	Sequence 6, Appli
433	1	10.9	5	US-08-459-871-6	Sequence 10, Appli
434	1	10.9	5	US-08-459-327-6	Sequence 26, Appl
435	2	10.9	5	US-08-911-321-10	Sequence 51, Appl
436	2	10.9	5	US-08-948-176-26	Sequence 6, Appli
437	2	10.9	5	US-08-928-292-51	Sequence 14, Appli
438	2	10.9	5	US-08-609-236-6	Sequence 16, Appli
439	2	10.9	5	US-08-494-151-14	Sequence 7, Appli
440	2	10.9	5	US-09-106-217-16	Sequence 8, Appli
441	4	10.9	5	US-09-171-337A-7	Sequence 2, Appli
442	2	10.9	5	US-09-171-337A-8	Sequence 2, Appli
443	4	10.9	5	US-08-588-113-2	Sequence 3, Appli
444	3	10.9	5	US-09-106-217-2	Sequence 1, Appl
445	3	10.9	5	US-08-840-204-3	Sequence 18, Appli
446	4	10.9	5	US-09-193-266-1	Sequence 19, Appl
447	1	10.9	5	US-08-073-807A-18	Sequence 5, Appli
448	4	10.9	5	US-08-213-419B-19	Sequence 5, Appli
449	1	10.9	5	US-08-036-210-5	Sequence 6, Appli
450	1	10.9	5	US-08-449-609-5	Sequence 6, Appli
451	1	10.9	5	US-08-121-714-6	Sequence 6, Appli
452	2	10.9	5	US-08-477-108A-6	Sequence 7, Appli
453	5	10.9	5	US-08-477-112-6	Sequence 7, Appli
454	2	10.9	5	PCT-US93-08322-6	Sequence 1, Appli
455	1	10.9	5	US-08-799-173A-7	Sequence 3, Appli
456	2	10.9	5	US-08-886-152-1	Sequence 23, Appl
457	2	10.9	5	US-08-886-152-3	Sequence 4, Appli
458	4	10.9	5	US-08-530-290-23	Sequence 4, Appli
459	1	10.9	5	US-09-144-914-4	Sequence 4, Appli
460	1	10.9	5	US-08-434-255-2	Sequence 4, Appli
461	1	10.9	5	US-08-434-255-4	Sequence 4, Appli
462	1	10.9	5	US-08-459-967-2	Sequence 2, Appli
463	1	10.9	5	US-08-459-967-4	Sequence 2, Appli
464	1	10.9	5	US-08-460-327-2	Sequence 2, Appli
465	1	10.9	5	US-08-459-871-2	Sequence 2, Appli

No.			Ref	C	Patent / US	Sequence
466	5	10.9	397	1	US-08-459-871-4	Sequence 4, Appli
467	5	10.9	398	4	US-08-810-712-4	Sequence 4, Appli
468	5	10.9	399	1	US-08-530-950-10	Sequence 10, Appl
469	5	10.9	399	2	US-08-874-186-92	Sequence 92, Appl
470	5	10.9	399	4	US-08-888-429A-10	Sequence 10, Appl
471	5	10.9	399	4	US-09-149-879-10	Sequence 10, Appl
472	5	10.9	399	2	US-09-553-498-2	Sequence 2, Appl
473	5	10.9	402	1	US-08-315-461-7	Sequence 7, Appli
474	5	10.9	402	3	US-08-948-997-4	Sequence 4, Appli
475	5	10.9	402	4	US-09-348-817A-4	Sequence 4, Appli
476	5	10.9	402	1	US-09-026-408-3	Sequence 3, Appli
477	5	10.9	409	4	US-09-640-305-6	Sequence 6, Appli
478	5	10.9	409	1	US-08-360-673-6	Sequence 6, Appli
479	5	10.9	415	4	US-09-100-193-2	Sequence 2, Appli
480	5	10.9	416	4	US-09-025-769B-280	Sequence 280, App
481	5	10.9	416	3	US-08-073-807A-2	Sequence 2, Appli
482	5	10.9	416	3	US-08-858-876A-4	Sequence 4, Appli
483	5	10.9	417	3	US-08-705-771-18	Sequence 18, Appl
484	5	10.9	424	3	US-08-247-908A-11	Sequence 11, Appl
485	5	10.9	424	5	US-08-453-942-11	Sequence 11, Appl
486	5	10.9	424	2	US-08-926-885A-11	Sequence 11, Appl
487	5	10.9	424	5	PCT-US94-05290-11	Sequence 11, Appl
488	5	10.9	428	3	US-09-118-319-5	Sequence 5, Appl
489	5	10.9	430	1	US-08-785-076-3	Sequence 3, Appli
490	5	10.9	432	1	US-08-785-076-2	Sequence 2, Appli
491	5	10.9	432	2	US-08-644-034A-2	Sequence 2, Appli
492	5	10.9	434	3	US-08-844-154-2	Sequence 5, Appli
493	5	10.9	434	3	US-09-126-192A-2	Sequence 5, Appli
494	5	10.9	436	5	US-08-716-873-5	Sequence 5, Appli
495	5	10.9	436	4	US-08-368-431-5	Sequence 10, Appl
496	5	10.9	437	1	US-08-764-543-1	Sequence 10, Appl
497	5	10.9	437	1	US-08-810-116-10	Sequence 36, Appl
498	5	10.9	437	2	US-07-930-548A-10	Sequence 36, Appl
499	5	10.9	437	6	US-08-677-049-9	Sequence 21, Appl
500	5	10.9	438	3	US-08-989-925-3	Sequence 6, Appli
501	5	10.9	438	4	US-09-167-049-9	Sequence 6, Appli
502	5	10.9	439	5	US-09-167-299-5	Sequence 10, Appli
503	5	10.9	439	4	US-08-716-873-2	Sequence 26, Appl
504	5	10.9	439	4	US-08-716-873-4	Sequence 51, Appl
505	5	10.9	439	4	US-09-368-431-2	Sequence 6, Appli
506	5	10.9	442	4	US-09-368-431-4	Sequence 14, Appl
507	5	10.9	442	1	US-08-363-255-4	Sequence 16, Appl
508	5	10.9	442	1	US-08-363-255-11	Sequence 12, Appli
509	5	10.9	444	4	US-09-252-292c-27	Sequence 7, Appli
510	5	10.9	445	1	US-08-353-400-33	Sequence 8, Appli
511	5	10.9	445	2	US-08-363-255-5	Patent No. 5455030
512	5	10.9	445	5	US-08-363-255-12	Sequence 3, Appli
513	5	10.9	447	2	US-08-902-585-2	Sequence 7, Appli
514	5	10.9	447	6	5455030-1	Sequence 41, Appl
515	5	10.9	448	3	US-08-570-157-3	Sequence 35, Appl
516	5	10.9	449	3	US-08-987-743-7	Sequence 5, Appli
517	5	10.9	450	4	US-09-268-992-41	Sequence 2, Appli
518	5	10.9	451	2	US-09-357-251-35	Sequence 11, Appl
519	5	10.9	452	1	US-08-290-978A-5	Sequence 26, Appl
520	5	10.9	452	2	US-08-780-869-5	Sequence 27, Appl
521	5	10.9	459	2	US-09-491-785-2	Sequence 36, Appl
522	5	10.9	461	2	US-08-463-587A-26	Sequence 5, Appli
523	5	10.9	461	4	US-08-463-667A-4	Sequence 33, Appl
524	5	10.9	461	5	US-08-923-854-26	Sequence 12, Appli
525	5	10.9	461	5	PCT-US91-0133-27	Sequence 2, Appli
526	5	10.9	464	4	US-08-353-400-36	Sequence 39, Appli
527	5	10.9	464	2	US-08-553-619B-9	Sequence 6, Appli
528	5	10.9	464	3	US-09-056-783-2	Sequence 7, Appli
529	5	10.9	466	4	US-09-268-992-39	Sequence 2, Appli
530	5	10.9	467	1	US-09-086-483A-6	Sequence 2, Appli
531	5	10.9	467	4	US-08-495-484-12	Sequence 4, Appli
532	5	10.9	468	1	US-08-303-569B-7	Sequence 2, Appli
533	5	10.9	468	2	US-08-116-247-7	Sequence 2, Appli
534	5	10.9	468	4	US-09-013-895A-2	Sequence 1, Appli
535	5	10.9	469	2	US-08-968-751-2	
536	5	10.9	469	4	US-09-052-089A-1	
537	5	10.9	469	4	US-08-749-903-1	
538	5	10.9	472	1		

Patent No. 5455030

No.	Ref	n	Score	Len	Publication	Sequence
539	472	1	10.9	5	US-08-749-903-3	Sequence 3, Appli
540	472	4	10.9	5	US-09-088-641-1	Sequence 1, Appli
541	472	4	10.9	5	US-09-088-641-3	Sequence 3, Appli
542	473	4	10.9	5	US-08-857-076-99	Sequence 99, Appl
543	473	4	10.9	5	US-09-171-945-131	Sequence 131, App
544	477	2	10.9	5	US-08-432-016-3	Sequence 3, Appli
545	477	2	10.9	5	US-08-684-594-3	Sequence 3, Appli
546	479	1	10.9	5	US-08-484-105-10	Sequence 10, Appl
547	479	1	10.9	5	US-08-484-106-10	Sequence 10, Appl
548	479	2	10.9	5	US-08-126-109-8	Sequence 8, Appli
549	480	2	10.9	5	US-08-828-488-8	Sequence 8, Appli
550	480	3	10.9	5	US-09-189-035-5	Sequence 5, Appli
551	480	3	10.9	5	US-09-382-086-5	Sequence 5, Appli
552	482	1	10.9	5	US-08-358-160-5	Sequence 5, Appli
553	482	3	10.9	5	US-09-135-639-2	Sequence 2, Appli
554	484	1	10.9	5	US-08-358-160-7	Sequence 7, Appli
555	492	3	10.9	5	US-08-724-466B-2	Sequence 2, Appli
556	492	3	10.9	5	US-08-882-164D-2	Sequence 2, Appli
557	493	1	10.9	5	US-07-615-448A-7	Sequence 7, Appli
558	493	2	10.9	5	US-08-196-361-7	Sequence 7, Appli
559	493	2	10.9	5	US-08-446-934-7	Sequence 7, Appli
560	493	2	10.9	5	US-08-448-128-7	Sequence 7, Appli
561	493	3	10.9	5	US-08-948-703-7	Sequence 7, Appli
562	497	3	10.9	5	US-09-047-125-3	Sequence 3, Appli
563	497	3	10.9	5	US-07-736-335E-3	Sequence 3, Appli
564	498	4	10.9	5	US-09-058-260-16	Sequence 16, Appl
565	501	1	10.9	5	US-09-149-105-15	Sequence 15, Appl
566	501	1	10.9	5	US-08-149-105-17	Sequence 17, Appl
567	501	1	10.9	5	US-08-317-847-15	Sequence 15, Appl
568	501	1	10.9	5	US-08-317-847-17	Sequence 17, Appl
569	503	2	10.9	5	US-08-781-802-2	Sequence 2, Appli
570	503	2	10.9	5	US-08-781-802-10	Sequence 10, Appl
571	503	2	10.9	5	US-08-781-802-12	Sequence 12, Appl
572	503	4	10.9	5	US-09-382-256-10	Sequence 10, Appl
573	503	4	10.9	5	US-08-694-078-2	Sequence 2, Appli
574	503	4	10.9	5	US-09-058-260-2	Sequence 2, Appli
575	503	4	10.9	5	US-09-058-260-10	Sequence 10, Appl
576	503	4	10.9	5	US-09-058-260-12	Sequence 12, Appl
577	503	4	10.9	5	US-09-058-260-28	Sequence 28, Appl
578	503	4	10.9	5	US-09-058-260-30	Sequence 30, Appl
579	503	4	10.9	5	US-09-395-115-10	Sequence 10, Appl
580	503	4	10.9	5	US-08-123-934A-10	Sequence 10, Appl
581	503	4	10.9	5	US-08-436-265-10	Sequence 10, Appl
582	503	5	10.9	5	PCT-US94-10080-10	Sequence 10, Appl
583	505	4	10.9	5	US-09-382-256-8	Sequence 8, Appli
584	505	4	10.9	5	US-09-382-256-16	Sequence 16, Appl
585	505	4	10.9	5	US-09-395-115-8	Sequence 8, Appli
586	505	4	10.9	5	US-09-395-115-16	Sequence 16, Appl
587	505	4	10.9	5	US-08-123-934A-8	Sequence 8, Appli
588	505	4	10.9	5	US-08-436-265-8	Sequence 8, Appli
589	505	5	10.9	5	US-08-436-265-16	Sequence 16, Appl
590	505	5	10.9	5	PCT-US94-10080-8	Sequence 8, Appli
591	508	2	10.9	5	US-08-724-281-1	Sequence 1, Appli
592	513	1	10.9	5	US-08-390-162-6	Sequence 6, Appli
593	513	1	10.9	5	US-08-685-945B-6	Sequence 6, Appli
594	513	3	10.9	5	US-08-764-870-7	Sequence 7, Appli
595	516	4	10.9	5	US-08-980-115-7	Sequence 7, Appli
596	524	1	10.9	5	US-08-927-219-140	Sequence 140, App
597	524	4	10.9	5	US-09-201-641-6	Sequence 6, Appli
598	525	2	10.9	5	US-08-624-125-2	Sequence 2, Appli
599	525	4	10.9	5	US-08-624-125-21	Sequence 21, Appl
600	527	4	10.9	5	US-08-818-024-4	Sequence 4, Appli
601	532	1	10.9	5	US-08-285-440-5	Sequence 5, Appli
602	532	4	10.9	5	US-08-630-349-5	Sequence 5, Appli
603	533	1	10.9	5	US-07-952-800-2	Sequence 2, Appli
604	533	1	10.9	5	US-08-770-544-4	Sequence 4, Appli
605	535	2	10.9	5	US-08-813-574-2	Sequence 2, Appli
606	539	4	10.9	5	US-08-818-024-4	Sequence 4, Appli
607	539	2	10.9	5	US-09-311-924-4	Sequence 4, Appli
608	539	2	10.9	5	US-09-334-775A-4	Sequence 4, Appli
609	541	4	10.9	5	US-08-789-275-6	Sequence 6, Appli
610	541	4	10.9	5	US-08-627-376-11	Sequence 11, Appl
611	542	3	10.9	5	US-08-968-563-16	Sequence 16, Appl
612	542	4	10.9	5	US-08-969-683A-16	Sequence 16, Appl
613	542	4	10.9	5	US-09-297-928-12	Sequence 12, Appl
614	543	2	10.9	5	US-08-469-412A-7	Sequence 7, Appli
615	543	4	10.9	5	US-08-697-610-2	Sequence 2, Appli
616	543	4	10.9	5	US-09-021-715-7	Sequence 7, Appli
617	543	4	10.9	5	US-08-349-357-2	Sequence 2, Appli
618	545	1	10.9	5	US-08-133-347-2	Sequence 2, Appli
619	545	4	10.9	5	US-08-500-635A-12	Sequence 12, Appl
620	549	4	10.9	5	US-09-167-151-12	Sequence 12, Appl
621	550	4	10.9	5	US-08-143-219-25	Sequence 25, Appl
622	551	2	10.9	5	US-08-436-771-9	Sequence 9, Appli
623	551	2	10.9	5	US-08-434-998-9	Sequence 9, Appli
624	551	2	10.9	5	US-08-487-797-9	Sequence 9, Appli
625	551	2	10.9	5	US-09-067-351-2	Sequence 2, Appli
626	551	5	10.9	5	US-09-360-490-2	Sequence 2, Appli
627	558	1	10.9	5	PCT-US95-02058-9	Sequence 9, Appli
628	558	4	10.9	5	US-08-285-440-6	Sequence 6, Appli
629	563	4	10.9	5	US-08-630-349-6	Sequence 6, Appli
630	567	4	10.9	5	US-09-311-924-2	Sequence 2, Appli
631	574	2	10.9	5	US-08-927-219-131	Sequence 131, App
632	577	1	10.9	5	US-08-836-854-21	Sequence 21, Appl
633	577	1	10.9	5	US-09-463-296-1	Sequence 1, Appli
634	580	1	10.9	5	US-08-484-105-24	Sequence 24, Appl
635	580	5	10.9	5	US-08-484-106-24	Sequence 24, Appl
636	581	3	10.9	5	US-08-309-512-6	Sequence 6, Appli
637	584	1	10.9	5	US-08-591-079-2	Sequence 2, Appli
638	584	5	10.9	5	PCT-US92-08756A-6	Sequence 6, Appli
639	585	1	10.9	5	US-08-619-812-6	Sequence 6, Appli
640	585	2	10.9	5	US-08-261-822A-8	Sequence 8, Appli
641	585	1	10.9	5	PCT-US95-07744A-8	Sequence 8, Appli
642	599	3	10.9	5	US-08-477-674-10	Sequence 10, Appl
643	600	6	10.9	5	US-08-473-791-10	Sequence 10, Appl
644	612	2	10.9	5	US-08-316-714-10	Sequence 10, Appl
645	612	2	10.9	5	US-08-473-673-10	Sequence 10, Appl
646	613	2	10.9	5	US-08-442-542-18	Sequence 18, Appl
647	613	3	10.9	5	US-08-765-469-18	Sequence 18, Appl
648	600	6	10.9	5	540706-1	Patent No. 5240706
649	612	2	10.9	5	US-08-746-283-31	Sequence 31, Appl
650	612	2	10.9	5	US-08-746-257A-29	Sequence 29, Appl
651	613	2	10.9	5	US-08-746-283-1	Sequence 1, Appli
652	613	2	10.9	5	US-08-746-283-3	Sequence 3, Appli
653	613	3	10.9	5	US-08-746-283-5	Sequence 5, Appli
654	615	2	10.9	5	US-08-746-257A-1	Sequence 4, Appli
655	623	4	10.9	5	US-08-860-091A-4	Sequence 4, Appli
656	626	2	10.9	5	US-09-104-068-4	Sequence 7, Appli
657	626	2	10.9	5	US-08-596-300A-7	Sequence 14, Appl
658	630	4	10.9	5	US-08-596-300A-14	Sequence 2, Appli
659	637	2	10.9	5	US-09-704-449-2	Sequence 2, Appli
660	638	4	10.9	5	US-09-104-068-2	Sequence 74, Appl
661	639	1	10.9	5	US-09-228-986-74	Sequence 2, Appli
662	639	1	10.9	5	US-08-466-390-2	Sequence 2, Appli
663	639	2	10.9	5	US-08-470-950-2	Sequence 2, Appli
664	641	1	10.9	5	US-08-467-781-2	Sequence 2, Appli
665	655	1	10.9	5	US-08-195-487-2	Sequence 2, Appli
666	670	4	10.9	5	US-08-483-924-2	Sequence 3, Appli
667	685	4	10.9	5	PCT-US93-06160-2	Sequence 3, Appli
668	685	2	10.9	5	US-07-718-535-3	Sequence 3, Appli
669	685	4	10.9	5	US-08-161-999-3	Sequence 25, Appl
670	686	4	10.9	5	US-09-178-252-25	Sequence 60, Appl
671	694	1	10.9	5	US-08-148-910-12	Sequence 12, Appl
672	694	5	10.9	5	US-08-448-937A-12	Sequence 12, Appl
673	706	4	10.9	5	US-08-980-080-2	Sequence 2, Appli
674	713	2	10.9	5	US-08-878-989-1	Sequence 1, Appli
675		4	10.9	5	US-09-136-282-2	Sequence 2, Appli
676		4	10.9	5	US-09-272-796-1	Sequence 2, Appli
677		1	10.9	5	US-09-423-439-60	Sequence 25, Appl
678		4	10.9	5	US-09-505-744-2	Sequence 2, Appli
679		2	10.9	5	US-08-768-301-4	Sequence 4, Appli
680		1	10.9	5	US-08-895-522-3	Sequence 3, Appli
681		3	10.9	5	US-09-195-391-3	Sequence 16, Appl
682		1	10.9	5	US-08-484-105-16	Sequence 16, Appl
683		1	10.9	5	US-08-484-106-16	Sequence 16, Appl
684		4	10.9	5	US-09-346-237-9	Sequence 9, Appli

#			US No.	Sequence
685	10.9	5	US-09-588-256-2	Sequence 2, Appli
686	10.9	5	US-08-840-236-1	Sequence 6, Appli
687	10.9	5	US-08-505-448A-1	Sequence 1, Appli
688	10.9	5	US-08-813-940-25	Sequence 25, Appl
689	10.9	5	US-09-129-075-4	Sequence 4, Appli
690	10.9	5	US-09-346-237-3	Sequence 3, Appli
691	10.9	5	US-09-346-237-13	Sequence 13, Appl
692	10.9	5	US-07-640-029-3	Sequence 4, Appli
693	10.9	5	US-07-921-807B-5	Sequence 5, Appli
694	10.9	5	US-08-441-944A-5	Sequence 5, Appli
695	10.9	5	US-08-439-992A-3	Sequence 3, Appli
696	10.9	5	US-07-640-029-4	Sequence 4, Appli
697	10.9	5	US-07-921-807B-6	Sequence 6, Appli
698	10.9	5	US-08-441-944A-6	Sequence 6, Appli
699	10.9	5	US-08-439-992A-4	Sequence 4, Appli
700	10.9	5	US-08-276-099A-16	Sequence 16, Appl
701	10.9	5	US-08-781-890-16	Sequence 16, Appl
702	10.9	5	US-08-674-887A-8	Sequence 8, Appli
703	10.9	5	US-09-010-928B-28	Sequence 28, Appl
704	10.9	5	US-08-951-844-8	Sequence 8, Appli
705	10.9	5	US-08-895-522-1	Sequence 1, Appli
706	10.9	5	US-09-195-391-1	Sequence 1, Appli
707	10.9	5	US-08-434-730-16	Sequence 16, Appl
708	10.9	5	US-09-413-814-32	Sequence 32, Appl
709	10.9	5	US-07-731-157A-7	Sequence 7, Appli
710	10.9	5	US-08-541-780-7	Sequence 7, Appli
711	10.9	5	US-09-012-710-11	Sequence 11, Appl
712	10.9	5	US-09-556-273-11	Sequence 11, Appl
713	10.9	5	US-08-232-538-15	Sequence 15, Appl
714	10.9	5	US-08-786-164-15	Sequence 15, Appl
715	10.9	5	US-09-012-710-10	Sequence 10, Appl
716	10.9	5	US-09-556-273-10	Sequence 10, Appl
717	10.9	5	US-08-366-276-2	Sequence 2, Appli
718	10.9	5	US-08-393-333-2	Sequence 2, Appli
719	10.9	5	US-08-909-954-4	Sequence 4, Appli
720	10.9	5	US-07-862-021B-10	Sequence 10, Appl
721	10.9	5	US-08-313-288B-10	Sequence 10, Appl
722	10.9	5	PCT-US93-03164-10	Sequence 10, Appl
723	10.9	5	US-08-629-291A-33	Sequence 33, Appl
724	10.9	5	US-08-658-335B-33	Sequence 33, Appl
725	10.9	5	US-09-199-637A-93	Sequence 93, Appl
726	10.9	5	US-08-233-788A-42	Sequence 42, Appl
727	10.9	5	US-07-640-029-2	Sequence 2, Appli
728	10.9	5	US-07-921-807B-3	Sequence 3, Appli
729	10.9	5	US-08-441-944A-3	Sequence 3, Appli
730	10.9	5	US-08-166-717D-6	Sequence 6, Appli
731	10.9	5	US-08-380-182-23	Sequence 23, Appl
732	10.9	5	US-08-439-992A-1	Sequence 1, Appli
733	10.9	5	US-07-997-133-1	Sequence 1, Appli
734	10.9	5	US-07-921-807B-4	Sequence 4, Appli
735	10.9	5	US-08-459-296-2	Sequence 2, Appli
736	10.9	5	US-09-010-928B-2	Sequence 2, Appli
737	10.9	5	US-08-441-944A-4	Sequence 4, Appli
738	10.9	5	US-08-451-822A-12	Sequence 12, Appl
739	10.9	5	US-08-439-992A-2	Sequence 2, Appli
740	10.9	5	US-08-323-430-12	Sequence 12, Appl
741	10.9	5	US-08-431-560-1	Sequence 1, Appli
742	10.9	5	US-08-413-118-3	Sequence 3, Appli
743	10.9	5	US-09-291-839-2	Sequence 2, Appli
744	10.9	5	US-08-413-118-106	Sequence 106, Appl
745	10.9	5	US-08-699-103B-10	Sequence 10, Appl
746	10.9	5	US-09-229-059-10	Sequence 10, Appl
747	10.9	5	US-08-395-580-2	Sequence 2, Appli
748	10.9	5	PCT-US95-02792-2	Sequence 2, Appli
749	10.9	5	US-08-785-241-4	Sequence 4, Appli
750	10.9	5	US-09-010-928B-2	Sequence 2, Appli
751	10.9	5	US-08-220-151-2	Sequence 2, Appli
752	10.9	5	US-08-220-151-3	Sequence 3, Appli
753	10.9	5	US-08-413-118-3	Sequence 3, Appli
754	10.9	5	US-08-413-118-106	Sequence 106, App
755	10.9	5	US-08-473-446-3	Sequence 3, Appli
756	10.9	5	US-08-473-446-3	Sequence 3, Appli
757	10.9	5	US-08-473-446-106	Sequence 106, App

#			pos	n	US No.	Sequence
758	10.9	5	888	2	US-08-861-464-6	Sequence 6, Appli
759	10.9	5	888	2	US-08-396-001-6	Sequence 6, Appli
760	10.9	5	888	4	US-09-323-433A-6	Sequence 6, Appli
761	10.9	5	903	2	US-08-209-521-24	Sequence 24, Appl
762	10.9	5	903	4	US-08-961-810-134	Sequence 134, App
763	10.9	5	903	4	US-08-352-902D-134	Sequence 134, App
764	10.9	5	903	4	US-09-228-986-78	Sequence 78, Appl
765	10.9	5	907	2	US-09-010-928B-4	Sequence 12, Appl
766	10.9	5	908	3	US-08-699-103B-12	Sequence 12, Appl
767	10.9	5	908	4	US-09-229-059-12	Sequence 9, Appli
768	10.9	5	921	4	US-09-206-800-9	Sequence 204, App
769	10.9	5	943	5	US-09-056-556-204	Sequence 14, Appl
770	10.9	5	948	1	US-08-602-228-14	Sequence 14, Appl
771	10.9	5	948	2	US-08-533-901B-14	Sequence 14, Appl
772	10.9	5	948	2	US-08-839-032A-14	Sequence 14, Appl
773	10.9	5	948	2	US-08-839-031A-14	Sequence 14, Appl
774	10.9	5	948	5	PCT-US95-12724-14	Sequence 5, Appli
775	10.9	5	956	1	US-08-897-443-3	Sequence 3, Appli
776	10.9	5	966	1	US-08-571-758-2	Sequence 2, Appli
777	10.9	5	966	1	US-08-909-984A-2	Sequence 2, Appli
778	10.9	5	966	1	US-08-909-983-2	Sequence 3, Appli
779	10.9	5	972	3	US-08-750-141A-2	Sequence 3, Appli
780	10.9	5	984	2	US-08-257-073-3	Sequence 120, App
781	10.9	5	984	2	US-08-184-009-120	Sequence 120, App
782	10.9	5	984	2	US-08-458-356-120	Sequence 120, App
783	10.9	5	984	2	US-08-460-736-120	Sequence 2, Appli
784	10.9	5	989	4	US-08-213-419B-2	Sequence 4, Appli
785	10.9	5	989	4	US-08-213-419B-4	Sequence 4, Appli
786	10.9	5	997	2	US-08-387-942C-4	Sequence 49, Appl
787	10.9	5	1050	3	US-09-045-632-49	Sequence 50, Appl
788	10.9	5	1050	3	US-09-045-632-50	Sequence 24, Appl
789	10.9	5	1066	4	US-09-004-838-24	Sequence 2, Appli
790	10.9	5	1115	3	US-08-323-477-2	Sequence 1, Appli
791	10.9	5	1123	4	US-09-408-865-1	Sequence 11, Appl
792	10.9	5	1127	3	US-09-150-460B-11	Sequence 4, Appli
793	10.9	5	1139	4	US-08-537-210A-4	Sequence 4, Appli
794	10.9	5	1139	4	US-09-113-825-4	Sequence 54, Appl
795	10.9	5	1141	1	US-08-131-365B-54	Sequence 54, Appl
796	10.9	5	1141	2	US-08-668-123-54	Sequence 6, Appli
797	10.9	5	1146	2	US-08-914-999-6	Sequence 38, Appl
798	10.9	5	1147	1	US-08-131-365B-38	Sequence 38, Appl
799	10.9	5	1147	2	US-08-668-123-38	Sequence 1, Appli
800	10.9	5	1164	3	US-08-589-756-1	Sequence 1, Appli
801	10.9	5	1164	4	US-09-206-800-1	Sequence 1, Appli
802	10.9	5	1164	1	US-09-206-898-1	Sequence 23, Appl
803	10.9	5	1186	4	US-09-178-252-23	Sequence 4, Appli
804	10.9	5	1203	2	US-09-075-272-4	Sequence 119, App
805	10.9	5	1222	4	US-09-004-838-119	Sequence 8, Appli
806	10.9	5	1227	1	US-08-448-170-8	Sequence 9, Appli
807	10.9	5	1227	4	US-08-961-803-9	Sequence 8, Appli
808	10.9	5	1245	2	US-08-158-232-8	Sequence 8, Appli
809	10.9	5	1245	1	US-08-304-626-8	Sequence 8, Appli
810	10.9	5	1245	2	US-08-611-928-8	Sequence 8, Appli
811	10.9	5	1248	3	US-09-173-891-8	Sequence 16, Appl
812	10.9	5	1257	3	US-08-726-214-16	Sequence 2, Appli
813	10.9	5	1275	2	US-08-750-152A-2	Sequence 2, Appli
814	10.9	5	1275	2	US-09-120-513-2	Sequence 1, Appli
815	10.9	5	1279	2	US-09-450-105-2	Sequence 2, Appli
816	10.9	5	1280	2	US-08-784-649A-2	Sequence 2, Appli
817	10.9	5	1280	2	US-08-583-276-19	Sequence 19, Appl
818	10.9	5	1280	1	US-08-752-447-2	Sequence 2, Appli
819	10.9	5	1280	6	5206352	Patent No. 5206352
820	10.9	5	1290	2	US-08-138-641-2	Sequence 2, Appli
821	10.9	5	1290	1	US-08-138-133-2	Sequence 2, Appli
822	10.9	5	1311	1	US-08-340-011-5	Sequence 5, Appli
823	10.9	5	1311	1	US-08-901-710-5	Sequence 5, Appli
824	10.9	5	1323	1	US-09-004-838-90	Sequence 90, Appl
825	10.9	5	1337	3	US-08-854-585-2	Sequence 2, Appli
826	10.9	5	1337	5	PCT-US95-05512-2	Sequence 2, Appli
827	10.9	5	1338	2	US-08-750-141A-3	Sequence 3, Appli
828	10.9	5	1356	1	US-08-810-116-8	Sequence 8, Appli
829	10.9	5	1356	2	US-07-930-548A-8	Sequence 8, Appli

#			ID		Patent No.	Sequence
831	10.9	5	1356	4	US-09-098-707A-2	Sequence 2, Appli
832	10.9	5	1356	4	US-09-770-170-6	Sequence 6, Appli
833	10.9	5	1356	4	US-09-483-539-2	Sequence 2, Appl
834	10.9	5	1362	3	US-08-874-678-33	Sequence 33, Appl
835	10.9	5	1362	3	US-08-643-839-33	Sequence 33, Appl
836	10.9	5	1363	1	US-08-425-061-23	Sequence 23, Appl
837	10.9	5	1363	4	US-08-825-886-23	Sequence 23, Appl
838	10.9	5	1426	4	US-09-136-574A-43	Sequence 43, Appl
839	10.9	5	1434	3	US-08-540-406-10	Sequence 10, Appl
840	10.9	5	1434	4	US-08-656-055-10	Sequence 10, Appl
841	10.9	5	1434	4	US-08-954-668-10	Sequence 10, Appl
842	10.9	5	1437	3	PCT-US95-13233-10	Sequence 10, Appl
843	10.9	5	1453	3	US-09-061-400-2	Sequence 2, Appli
844	10.9	5	1453	2	US-09-001-273-2	Sequence 2, Appli
845	10.9	5	1453	2	US-08-843-459A-2	Sequence 2, Appli
846	10.9	5	1463	1	US-08-220-603A-11	Sequence 11, Appl
847	10.9	5	1494	3	US-08-755-587-186	Sequence 186, App
848	10.9	5	1572	2	US-08-290-731C-5	Sequence 5, Appli
849	10.9	5	1581	4	US-09-110-517-2	Sequence 2, Appli
850	10.9	5	1596	3	US-09-356-952-3	Sequence 3, Appli
851	10.9	5	1627	2	US-07-665-792E-9	Sequence 9, Appli
852	10.9	5	1786	2	US-08-477-451-16	Sequence 16, Appl
853	10.9	5	1786	4	US-08-973-462-8	Sequence 8, Appli
854	10.9	5	1817	4	US-09-004-838-125	Sequence 125, App
855	10.9	5	1852	1	US-08-425-061-24	Sequence 24, Appl
856	10.9	5	1852	2	US-08-825-886-24	Sequence 24, Appl
857	10.9	5	1863	1	US-08-425-061-16	Sequence 16, Appl
858	10.9	5	1863	1	US-08-598-591-2	Sequence 2, Appli
859	10.9	5	1863	1	US-08-480-784-2	Sequence 2, Appli
860	10.9	5	1863	1	US-08-483-553-2	Sequence 2, Appli
861	10.9	5	1863	1	US-08-487-002-2	Sequence 2, Appli
862	10.9	5	1863	1	US-08-483-554B-2	Sequence 2, Appli
863	10.9	5	1863	1	US-08-798-691-2	Sequence 2, Appli
864	10.9	5	1863	1	US-08-798-691-4	Sequence 4, Appli
865	10.9	5	1863	1	US-08-798-691-6	Sequence 6, Appli
866	10.9	5	1863	2	US-08-488-011B-2	Sequence 2, Appli
867	10.9	5	1863	2	US-08-825-886-16	Sequence 16, Appl
868	10.9	5	1863	2	US-08-603-753D-2	Sequence 2, Appli
869	10.9	5	1863	3	US-08-825-487A-2	Sequence 2, Appli
870	10.9	5	1863	3	US-08-825-487A-4	Sequence 4, Appli
871	10.9	5	1863	3	US-08-825-487A-6	Sequence 6, Appli
872	10.9	5	1863	4	US-09-074-476-2	Sequence 2, Appli
873	10.9	5	1863	4	US-09-074-476-4	Sequence 4, Appli
874	10.9	5	1863	4	US-09-074-476-6	Sequence 6, Appli
875	10.9	5	1863	4	US-09-099-753-2	Sequence 2, Appli
876	10.9	5	1863	4	US-08-850-727-2	Sequence 2, Appli
877	10.9	5	1863	2	US-08-986-106-2	Sequence 2, Appli
878	10.9	5	1863	1	US-09-007-678B-49	Sequence 49, Appl
879	10.9	5	1863	2	PCT-US95-10202-2	Sequence 2, Appli
880	10.9	5	1863	5	PCT-US95-10203-2	Sequence 2, Appli
881	10.9	5	1863	5	PCT-US95-10220-2	Sequence 2, Appli
882	10.9	5	1872	1	US-08-188-582-14	Sequence 14, Appl
883	10.9	5	1872	1	US-08-188-582-11	Sequence 11, Appl
884	10.9	5	1872	6	536025-6	Patent No. 536025
885	10.9	5	1873	1	US-08-435-675B-4	Sequence 4, Appli
886	10.9	5	1873	1	US-08-336-257A-7	Sequence 7, Appli
887	10.9	5	1893	1	US-08-188-582-11	Sequence 11, Appl
888	10.9	5	1893	2	US-08-646-715-11	Sequence 11, Appl
889	10.9	5	1940	2	US-08-644-271-30	Sequence 30, Appl
890	10.9	5	1964	2	US-08-790-912-3	Sequence 3, Appli
891	10.9	5	2052	2	US-08-790-912-2	Sequence 2, Appli
892	10.9	5	2231	1	US-08-153-799-16	Sequence 16, Appl
893	10.9	5	2285	3	US-09-308-375-2	Sequence 2, Appli
894	10.9	5	2289	4	US-09-051-019-2	Sequence 2, Appli
895	10.9	5	2289	2	US-08-283-857-1	Sequence 1, Appli
896	10.9	5	2324	1	PCT-US95-09819-1	Sequence 1, Appli
897	10.9	5	2324	5	5455158-1	Patent No. 5455158
898	10.9	5	2327	6	US-09-016-366A-12	Sequence 12, Appl
899	10.9	5	2386	3	US-08-157-005-2	Sequence 2, Appli
900	10.9	5	2396	4	US-08-747-863-2	Sequence 2, Appli
901	10.9	5	2396	1	US-08-194-468-2	Sequence 2, Appli
902	10.9	5	2441	3	US-08-961-739-2	Sequence 2, Appli
903	10.9	5	2446	2	US-08-551-356-2	Sequence 2, Appli
904	10.9	5	2446	5	PCT-US93-12687-2	Sequence 2, Appli
905	10.9	5	2482	5	US-08-328-254-6	Sequence 6, Appl
906	10.9	5	2516	3	US-08-374-077C-2	Sequence 2, Appl
907	10.9	5	2516	4	US-08-895-590-2	Sequence 3, Appli
908	10.9	5	2627	2	US-08-751-189-3	Sequence 3, Appli
909	10.9	5	2627	2	US-09-060-836-3	Sequence 3, Appli
910	10.9	5	2627	4	US-09-184-445-3	Sequence 4, Appli
911	10.9	5	2629	2	US-08-751-189-4	Sequence 4, Appli
912	10.9	5	2629	2	US-09-060-836-4	Sequence 4, Appli
913	10.9	5	2629	1	US-09-184-445-4	Sequence 4, Appli
914	10.9	5	2703	5	US-08-185-432-19	Sequence 19, Appl
915	10.9	5	2713	5	PCT-US96-01735-1	Sequence 1, Appl
916	10.9	5	2763	2	US-08-496-944-2	Sequence 2, Appl
917	10.9	5	2802	4	US-09-542-331-1	Sequence 1, Appli
918	10.9	5	3072	4	US-09-413-814-93	Sequence 93, Appl
919	10.9	5	3079	4	US-09-413-814-80	Sequence 80, Appl
920	10.9	5	3118	2	US-08-457-273B-8	Sequence 8, Appli
921	10.9	5	3119	1	US-08-246-982A-16	Sequence 16, Appl
922	10.9	5	3119	1	US-08-53-265-16	Sequence 16, Appl
923	10.9	5	3144	1	US-08-246-982A-6	Sequence 6, Appli
924	10.9	5	3144	2	US-08-453-265-6	Sequence 6, Appli
925	10.9	5	3144	2	US-08-457-273B-42	Sequence 42, Appl
926	10.9	5	3144	3	US-08-556-419-21	Sequence 21, Appl
927	10.9	5	3144	4	US-09-041-886-15	Sequence 15, Appl
928	10.9	5	3174	2	US-08-477-451-3	Sequence 3, Appli
929	10.9	5	3248	5	US-08-353-700-1	Sequence 1, Appli
930	10.9	5	3248	5	PCT-US95-16216-1	Sequence 1, Appli
931	10.9	5	7257	3	US-09-335-409-5	Sequence 5, Appli
932	10.9	5	7257	4	US-09-568-102-5	Sequence 5, Appli
933	10.9	5	7257	4	US-09-567-969-5	Sequence 5, Appli
934	10.9	5	7257	4	US-09-568-480-5	Sequence 5, Appli
935	10.9	5	7257	4	US-09-568-486-5	Sequence 5, Appli
936	10.9	5	7257	4	US-09-568-472-5	Sequence 5, Appli
937	8.7	4	4	1	US-07-990-301A-10	Sequence 10, Appl
938	8.7	4	4	1	US-08-653-740-26	Sequence 26, Appl
939	8.7	4	4	2	US-08-482-651-5	Sequence 5, Appli
940	8.7	4	4	2	US-08-803-899-25	Sequence 25, Appl
941	8.7	4	4	2	US-09-073-594-26	Sequence 26, Appl
942	8.7	4	4	3	US-09-275-925-26	Sequence 26, Appl
943	8.7	4	4	4	US-08-435-568A-15	Sequence 15, Appl
944	8.7	4	4	4	US-08-602-999A-104	Sequence 104, App
945	8.7	4	4	4	US-09-113-921-90	Sequence 90, Appl
946	8.7	4	4	4	US-09-113-921-91	Sequence 91, Appl
947	8.7	4	4	5	US-08-660-092-8	Sequence 8, Appli
948	8.7	4	4	2	US-09-198-723A-21	Sequence 21, Appl
949	8.7	4	4	2	US-09-066-481-20	Sequence 20, Appl
950	8.7	4	4	2	US-08-918-288-80	Sequence 80, Appl
951	8.7	4	4	3	US-09-282-357-80	Sequence 80, Appl
952	8.7	4	4	4	US-08-278-865-104	Sequence 104, App
953	8.7	4	4	4	US-09-344-456-4	Sequence 4, Appli
954	8.7	4	5	1	US-07-956-848A-1	Sequence 1, Appli
955	8.7	4	5	1	US-07-956-848A-28	Sequence 28, Appl
956	8.7	4	5	1	US-07-956-848A-29	Sequence 29, Appl
957	8.7	4	5	1	US-09-095-726-71	Sequence 71, Appl
958	8.7	4	5	1	US-08-096-043-68	Sequence 68, Appl
959	8.7	4	5	1	US-08-096-023A-76	Sequence 76, Appl
960	8.7	4	5	1	US-08-451-947-75	Sequence 75, Appl
961	8.7	4	5	1	US-08-477-727A-54	Sequence 54, Appl
962	8.7	4	5	1	US-08-471-956-1	Sequence 1, Appli
963	8.7	4	5	1	US-08-471-956-28	Sequence 28, Appl
964	8.7	4	5	1	US-08-471-956-29	Sequence 29, Appl
965	8.7	4	5	2	US-08-424-826A-75	Sequence 75, Appl
966	8.7	4	5	2	US-08-347-335A-8	Sequence 8, Appli
967	8.7	4	5	2	US-08-441-513A-8	Sequence 8, Appli
968	8.7	4	5	3	US-08-751-767A-25	Sequence 25, Appl
969	8.7	4	5	3	US-08-751-767A-27	Sequence 27, Appl
970	8.7	4	5	3	US-08-928-694-75	Sequence 75, Appl
971	8.7	4	5	3	US-08-855-925A-7	Sequence 7, Appli
972	8.7	4	5	3	US-08-581-662-10	Sequence 10, Appl
973	8.7	4	5	3	US-08-581-662-28	Sequence 28, Appl
974	8.7	4	5	4	US-09-194-613-11	Sequence 11, Appl
975	8.7	4	5	4	US-09-187-859-2213	Sequence 2213, Ap
976	8.7	4	5	4	US-09-187-859-2543	Sequence 2543, Ap

```
977   4   8.7   5   5   PCT-US91-06950-75    Sequence 75, Appl
978   4   8.7   6   1   US-08-095-726-61     Sequence 61, Appl
979   4   8.7   6   1   US-08-095-726-65     Sequence 65, Appl
980   4   8.7   6   1   US-08-096-043-58     Sequence 58, Appl
981   4   8.7   6   1   US-08-096-043-62     Sequence 62, Appl
982   4   8.7   6   1   US-08-269-257-12     Sequence 12, Appl
983   4   8.7   6   1   US-08-264-002-15     Sequence 15, Appl
984   4   8.7   6   1   US-07-890-422B-9     Sequence 9, Appli
985   4   8.7   6   1   US-07-890-422B-10    Sequence 10, Appl
986   4   8.7   6   1   US-07-890-422B-13    Sequence 13, Appl
987   4   8.7   6   1   US-07-890-422B-14    Sequence 14, Appl
988   4   8.7   6   1   US-08-096-623A-66    Sequence 66, Appl
989   4   8.7   6   1   US-08-451-947-76     Sequence 70, Appl
990   4   8.7   6   1   US-08-451-947-76     Sequence 76, Appl
991   4   8.7   6   1   US-08-719-331-2      Sequence 2, Appli
992   4   8.7   6   2   US-08-724-548-50     Sequence 50, Appl
993   4   8.7   6   2   US-08-724-548-51     Sequence 51, Appl
994   4   8.7   6   2   US-08-424-826A-76    Sequence 76, Appl
995   4   8.7   6   2   US-08-463-667A-15    Sequence 15, Appl
996   4   8.7   6   2   US-08-465-380-18     Sequence 18, Appl
997   4   8.7   6   2   US-08-480-478-4      Sequence 4, Appli
998   4   8.7   6   2   US-08-486-399-18     Sequence 18, Appl
999   4   8.7   6   2   US-08-486-399-18     Sequence 18, Appl
1000  4   8.7   6   2   US-08-461-965-18     Sequence 18, Appl
```

ALIGNMENTS

RESULT 1
PCT-US93-03027-2
; Sequence 2, Application PC/TUS9303027
; GENERAL INFORMATION:
;   APPLICANT: LEONARD, WARREN; TOLEDANO,
;   APPLICANT: MICHEL
;   TITLE OF INVENTION: CONTROL AND/OR
;   TITLE OF INVENTION: PREVENTION OF BINDING OF NF- B/REL/DORSAL
;   TITLE OF INVENTION:
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: MORGAN & FINNEGAN
;   STREET: 345 PARK AVENUE
;   CITY: NEW YORK
;   STATE: NEW YORK
;   COUNTRY: USA
;   ZIP: 10154
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: FLOPPY DISK
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WORDPERFECT 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US93/03027
;   FILING DATE: 19930401
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/862,987
;   FILING DATE: 06-APR-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DOROTHY R. AUTH
;   REGISTRATION NUMBER: P-36,434
;   REFERENCE/DOCKET NUMBER: 2026-4010 PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-758-4800
;   TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 558
;   TYPE: AMINO ACID
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   ORGANISM: mouse
;   STRAIN:
;   INDIVIDUAL ISOLATE:
;   DEVELOPMENTAL STAGE:
;   HAPLOTYPE:
;   TISSUE TYPE:
;   CELL TYPE: fibroblast
;   CELL LINE: NIH-3T3
;   ORGANELE:
;   FEATURE:
;   NAME/KEY: RelB
;   LOCATION:
;   IDENTIFICATION METHOD: contains region
;   IDENTIFICATION METHOD: of high similarity with proteins of
;   IDENTIFICATION METHOD: Rel family.
;   OTHER INFORMATION:
;   PUBLICATION INFORMATION:
;   AUTHORS: Ryseck, R.-P.; Bull, P.;
;   AUTHORS: Takamiya, M.; Bours, V.; Siebenlist,
;   AUTHORS: U.; Dobrzanski, P.; Bravo, R.
;   TITLE: RelB, a new Rel family
;   TITLE: transcription activator that can
;   TITLE: interact with p50-NF- -B.
;   JOURNAL: Mol. Cell. Biol.
;   VOLUME: 12
;   ISSUE:
;   PAGES: 674-684
;   DATE: FEB-1992
;   DOCUMENT NUMBER:
;   PUBLICATION DATE:
;   FILING DATE:
;   RELEVANT RESIDUES IN SEQ ID NO:
PCT-US93-03027-2

Query Match   15.2%; Score 7; DB 5; Length 558;
Best Local Similarity   100.0%; Pred. No. 13;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   16 TDELEII 22
        |||||||
Db   33 TDELEII 39

RESULT 2
US-08-836-567-8
; Sequence 8, Application US/08836567
; Patent No. 6130367
; GENERAL INFORMATION:
;   APPLICANT: Kossmann, Jens
;   APPLICANT: Springer, Franziska
;   APPLICANT: Abel, Gernot
;   TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES
;   TITLE OF INVENTION: INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC
;   TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FISH & NEAVE
;   STREET: 1251 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10020
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/836,567
;   FILING DATE: 24-JUL-1997
;   CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/EP95/04415

```
        FILING DATE: 09-NOV-1995
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: DE P 44 41 408.0
        FILING DATE: 10-NOV-1994
    ATTORNEY/AGENT INFORMATION:
        NAME: Haley Jr., James F.
        REGISTRATION NUMBER: 27,794
        REFERENCE/DOCKET NUMBER: Agrevo-4
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 212-596-9000
        TELEFAX: 212-596-9090
    INFORMATION FOR SEQ ID NO: 8:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 767 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
        MOLECULE TYPE: protein
US-08-836-567-8

Query Match         15.2%;  Score 7;  DB 4;  Length 767;
Best Local Similarity  100.0%;  Pred. No. 17;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   24 GGGSLS 30
        ||||||
Db  108 GGGSLS 114

RESULT 3
US-08-370-567-18
; Sequence 18, Application US/08370567
; Patent No. 5655726
; GENERAL INFORMATION:
        APPLICANT: Rosenberg, Steven
        APPLICANT: Doyle, Michael
        APPLICANT: Goodson, Robert
        TITLE OF INVENTION: Peptide Inhibitors of Urokinase Receptor
        TITLE OF INVENTION: Activity
        NUMBER OF SEQUENCES: 37
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Chiron Corporation
        STREET: 4560 Horton Street
        CITY: Emeryville
        STATE: CA
        COUNTRY: USA
        ZIP: 94608
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/370,567
        FILING DATE:
        CLASSIFICATION: 530
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US/08/061,514
        FILING DATE:
    ATTORNEY/AGENT INFORMATION:
        NAME: Green, Grant D.
        REGISTRATION NUMBER: 31,259
        REFERENCE/DOCKET NUMBER: 0941.001
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 510-601-2706
        TELEFAX: 510-655-3542
    INFORMATION FOR SEQ ID NO: 18:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 17 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
        MOLECULE TYPE: peptide
        HYPOTHETICAL: NO
US-08-370-567-18


Query Match         13.0%;  Score 6;  DB 1;  Length 17;
Best Local Similarity  100.0%;  Pred. No. 5.2;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   13 EIKTDE 18
        ||||||
Db    2 EIKTDE 7

RESULT 4
US-08-438-759-18
; Sequence 18, Application US/08438759
; Patent No. 5679782
; GENERAL INFORMATION:
        APPLICANT: Rosenberg, Steven
        APPLICANT: Doyle, Michael
        APPLICANT: Goodson, Robert
        TITLE OF INVENTION: Peptide Inhibitors of Urokinase Receptor
        TITLE OF INVENTION: Activity
        NUMBER OF SEQUENCES: 37
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Chiron Corporation
        STREET: 4560 Horton Street
        CITY: Emeryville
        STATE: CA
        COUNTRY: USA
        ZIP: 94608
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC Compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/438,759
        FILING DATE:
        CLASSIFICATION: 536
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US/08/370,567
        FILING DATE:
        APPLICATION NUMBER: US/08/061,514
        FILING DATE:
    ATTORNEY/AGENT INFORMATION:
        NAME: Green, Grant D.
        REGISTRATION NUMBER: 31,259
        REFERENCE/DOCKET NUMBER: 0941.001
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 510-601-2706
        TELEFAX: 510-655-3542
    INFORMATION FOR SEQ ID NO: 18:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 17 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
        MOLECULE TYPE: peptide
        HYPOTHETICAL: NO
US-08-438-759-18


Query Match         13.0%;  Score 6;  DB 1;  Length 17;
Best Local Similarity  100.0%;  Pred. No. 5.2;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   13 EIKTDE 18
        ||||||
Db    2 EIKTDE 7

RESULT 5
PCT-US94-05684-18
; Sequence 18, Application PC/TUS9405684
; GENERAL INFORMATION:
        APPLICANT: Chiron Corporation
        TITLE OF INVENTION: Peptide Inhibitors of Urokinase Receptor
```

; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05684
; FILING DATE: 19 MAY 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 0941.100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2706
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
PCT-US94-05684-18

Query Match   13.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   13 EIKTDE 18
        ||||||
Db   2 EIKTDE 7

RESULT 6
US-08-652-558-52
; Sequence 52, Application US/08652558
; Patent No. 5861155
; GENERAL INFORMATION:
; APPLICANT: LIN, AUGUSTINE YEE-THARN
; TITLE OF INVENTION: HUMANIZED ANTIBODIES AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & WITCOFF
; STREET: 75 STATE STREET, 23RD FLOOR
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,558
; FILING DATE: JUNE 6, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/IB94/00387
; FILING DATE: NOVEMBER 21, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: YANKWICH, LEON R.
; REGISTRATION NUMBER: 30,237

; REFERENCE/DOCKET NUMBER: 95,497-L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-345-9100
; TELEFAX: 617-345-9111
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-652-558-52

Query Match   13.0%; Score 6; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 31;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   23 IGGSGS 28
        ||||||
Db   51 IGGSGS 56

RESULT 7
US-09-367-953B-120
; Sequence 120, Application US/09367953B
; Patent No. 6287572
; GENERAL INFORMATION:
; APPLICANT: Kingsman, Alan J
; APPLICANT: Kingsman, Susan M
; TITLE OF INVENTION: Anti-HIV peptides and proteins
; FILE REFERENCE: 550-154
; CURRENT APPLICATION NUMBER: US/09/367,953B
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: PCT/GB98/00563
; PRIOR FILING DATE: 1998-02-23
; PRIOR APPLICATION NUMBER: GB 9703802.0
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 120
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 2
US-09-367-953B-120

Query Match   13.0%; Score 6; DB 4; Length 134;
Best Local Similarity 100.0%; Pred. No. 37;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   15 KTDELE 20
        ||||||
Db   12 KTDELE 17

RESULT 8
US-09-367-953B-121
; Sequence 121, Application US/09367953B
; Patent No. 6287572
; GENERAL INFORMATION:
; APPLICANT: Kingsman, Alan J
; APPLICANT: Kingsman, Susan M
; TITLE OF INVENTION: Anti-HIV peptides and proteins
; FILE REFERENCE: 550-154
; CURRENT APPLICATION NUMBER: US/09/367,953B
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: PCT/GB98/00563
; PRIOR FILING DATE: 1998-02-23
; PRIOR APPLICATION NUMBER: GB 9703802.0
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 121
; LENGTH: 134

```
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 2
US-09-367-953B-121

Query Match    13.0%;  Score 6;  DB 4;  Length 134;
Best Local Similarity 100.0%;  Pred. No. 37;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    15  KTDELE  20
          ||||||
Db    12  KTDELE  17

RESULT 9
US-08-482-918-56
; Sequence 56, Application US/08482918
; Patent No. 6207417
; GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Stem Cell Factor
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/482,918
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/33005
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX: 25-3856
;   INFORMATION FOR SEQ ID NO: 56:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 282 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-482-918-56

Query Match    13.0%;  Score 6;  DB 4;  Length 282;
Best Local Similarity 100.0%;  Pred. No. 74;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    32  FFRLFN  37
          ||||||
Db   139  FFRLFN  144

RESULT 10
US-09-224-681-56
; Sequence 56, Application US/09224681
; Patent No. 6207454
; GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Method for Enhancing the Efficiency of Gene
;   TITLE OF INVENTION: Transfer with Stem Cell Factor (SCF) Polypeptide
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/224,681
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/005,893
;   FILING DATE: 12-JAN-1998
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/449,653
;   FILING DATE: 24-MAY-1995
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/982,255
;   FILING DATE: 25-NOV-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/589,701
;   FILING DATE: 01-OCT-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/573,616
;   FILING DATE: 24-AUG-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/537,198
;   FILING DATE: 11-JUN-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/422,383
;   FILING DATE: 16-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/35199
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 56:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 282 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-224-681-56

Query Match    13.0%;  Score 6;  DB 4;  Length 282;
Best Local Similarity 100.0%;  Pred. No. 74;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    32  FFRLFN  37
          ||||||
Db   139  FFRLFN  144

RESULT 11
US-08-336-728A-56
; Sequence 56, Application US/08336728A
```

```
  Patent No. 6207802
GENERAL INFORMATION:
  APPLICANT:  Zsebo, Krisztina M.
  APPLICANT:  Bosselman, Robert A.
  APPLICANT:  Suggs, Sidney V.
  APPLICANT:  Martin, Francis H.
  TITLE OF INVENTION:  Stem Cell Factor
  NUMBER OF SEQUENCES:  104
  CORRESPONDENCE ADDRESS:
    ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
    STREET:  6300 Sears Tower, 233 South Wacker Drive
    CITY:  Chicago
    STATE:  Illinois
    COUNTRY:  United States of America
    ZIP:  60606-6402
  COMPUTER READABLE FORM:
    MEDIUM TYPE:  Floppy disk
    COMPUTER:  IBM PC compatible
    OPERATING SYSTEM:  PC-DOS/MS-DOS
    SOFTWARE:  PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/336,728A
    FILING DATE:  09-NOV-1994
    CLASSIFICATION:  424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  07/982,255
    FILING DATE:  25-NOV-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  07/589,701
    FILING DATE:  01-OCT-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  07/573,616
    FILING DATE:  24-AUG-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  07/537,198
    FILING DATE:  11-JUN-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  07/422,383
    FILING DATE:  16-OCT-1989
  ATTORNEY/AGENT INFORMATION:
    NAME:  Clough, David W.
    REGISTRATION NUMBER:  36,107
    REFERENCE/DOCKET NUMBER:  01017/32956
  TELECOMMUNICATION INFORMATION:
    TELEPHONE:  312/474-6300
    TELEFAX:  312/474-0448
    TELEX:  25-3856
INFORMATION FOR SEQ ID NO:  56:
  SEQUENCE CHARACTERISTICS:
    LENGTH:  282 amino acids
    TYPE:  amino acid
    STRANDEDNESS:  single
    TOPOLOGY:  linear
    MOLECULE TYPE:  protein
US-08-336-728A-56

Query Match             13.0%;  Score 6;  DB 4;  Length 282;
Best Local Similarity  100.0%;  Pred. No. 74;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    32 FFRLFN 37
         ||||||
Db   139 FFRLFN 144

RESULT 12
US-09-383-586-33
; Sequence 33, Application US/09383586
; Patent No. 6242419
; GENERAL INFORMATION:
;   APPLICANT:  Strachan, Lorna
;   APPLICANT:  Sleeman, Matthew
;   APPLICANT:  Abernethy, Nevin
```

```
;   APPLICANT:  Onrust, Rene
;   APPLICANT:  Kumble, Anand
;   APPLICANT:  Murison, Greg
;   TITLE OF INVENTION:  Compounds isolated from stromal cells
;   TITLE OF INVENTION:  and methods for their use
;   FILE REFERENCE:  11000.1037c1
;   CURRENT APPLICATION NUMBER:  US/09/383,586
;   CURRENT FILING DATE:  1999-08-26
;   NUMBER OF SEQ ID NOS:  38
;   SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH:  322
;   TYPE:  PRT
;   ORGANISM:  Human
US-09-383-586-33

Query Match             13.0%;  Score 6;  DB 4;  Length 322;
Best Local Similarity  100.0%;  Pred. No. 84;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     4 TLSLKN 9
         ||||||
Db   203 TLSLKN 208

RESULT 13
US-09-079-415-2
; Sequence 2, Application US/09079415
; Patent No. 6013452
; GENERAL INFORMATION:
;   APPLICANT:  Christensen, Tove
;   APPLICANT:  Lehmbeck, Jan
;   TITLE OF INVENTION:  A Fungus Wherein The areA, pepC and/or
;   TITLE OF INVENTION:  pepE Genes Have Been Inactivated
;   NUMBER OF SEQUENCES:  12
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  No. 6013452disk of No. 6013452th America, Inc.
;     STREET:  405 Lexington Avenue
;     CITY:  New York
;     STATE:  NY
;     COUNTRY:  USA
;     ZIP:  10174
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/079,415
;     FILING DATE:  14-MAY-1998
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Rozek, Carol
;     REGISTRATION NUMBER:  36,993
;     REFERENCE/DOCKET NUMBER:  4657.204-US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  212-867-0123
;     TELEFAX:  212-878-9655
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  397 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-09-079-415-2

Query Match             13.0%;  Score 6;  DB 3;  Length 397;
Best Local Similarity  100.0%;  Pred. No. 1e+02;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    25 GSGSLS 30
         ||||||
Db   146 GSGSLS 151
```

```
RESULT 14
US-08-328-314-2
; Sequence 2, Application US/08328314
; Patent No. 5674728
; GENERAL INFORMATION:
;   APPLICANT: Buxton, Frank
;   APPLICANT: Jarai, Gabor
;   APPLICANT: Visser, Jacob
;   TITLE OF INVENTION: Fungal Protease
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;    ADDRESSEE: CIBA-GEIGY Corporation
;    STREET: 7 Skyline Drive
;    CITY: Hawthorne
;    STATE: New York
;    COUNTRY: USA
;    ZIP: 10532
;   COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: ASCII editor
;   CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/328,314
;    FILING DATE: TBA
;    CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;    NAME: Spruill, W. Murray
;    REGISTRATION NUMBER: 32,943
;    REFERENCE/DOCKET NUMBER: 4-19746/A
;   TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (919) 541-8615
;    TELEFAX: (919) 541-8689
;   INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 398 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-328-314-2

Query Match       13.0%; Score 6; DB 1; Length 398;
Best Local Similarity  100.0%; Pred. No. 1e+02;
Matches  6; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy   25 GSGSLS 30
        ||||||
Db  147 GSGSLS 152


RESULT 15
US-08-731-045-2
; Sequence 2, Application US/08731045
; Patent No. 5756338
; GENERAL INFORMATION:
;   APPLICANT: Buxton, Frank
;   APPLICANT: Jarai, Gabor
;   APPLICANT: Visser, Jacob
;   TITLE OF INVENTION: Fungal Protease
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;    ADDRESSEE: CIBA-GEIGY CORPORATION
;    STREET: 520 White Plains Road
;    CITY: Tarrytown
;    STATE: New York
;    COUNTRY: USA
;    ZIP: 10591-9725
;   COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: ASCII Editor
;   CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/731,045
;    FILING DATE: Herewith
;    CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: 08/328,314
;    FILING DATE: October 24, 1994
;   ATTORNEY/AGENT INFORMATION:
;    NAME: Ferraro, Gregory D.
;    REGISTRATION NUMBER: 36,134
;    REFERENCE/DOCKET NUMBER: 4-19746/A/DIV
;   TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (908) 277-3318
;    TELEFAX: (908) 277-4306
;   INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 398 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-731-045-2

Query Match       13.0%; Score 6; DB 1; Length 398;
Best Local Similarity  100.0%; Pred. No. 1e+02;
Matches  6; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy   25 GSGSLS 30
        ||||||
Db  147 GSGSLS 152

Search completed: November  5, 2002, 11:06:56
Job time : 38.1045 secs
```

```
                    GenCore version 5.1.3
              Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      November  5, 2002, 11:05:33 ; Search time 144.179 Seconds
                                           (without alignments)
                                           112.299 Million cell updates/sec

Title:            US-09-833-017B-2
Perfect score:    46
Sequence:         1 MKKTLSLKNDFKEIKTDELE........GSLSTFFRLFNRSFTQALGK 46

Scoring table:    OLIGO
                  Gapop 60.0 , Gapext 60.0

Searched:         3502263 seqs, 351980561 residues

Word size :       0

Total number of hits satisfying chosen parameters:       3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :       Pending_Patents_AA_Main:*
                 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep:*
                 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
                 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
                 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
                 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
                 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
                 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
                 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
                 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
                10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
                11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
                12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
                13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
                14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
                15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
                16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
                17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
                18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
                19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
                20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
                21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
                22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
                23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
                24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
                25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
                26: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	46	22	US-09-833-017-2	Sequence 2, Appli
2	46	100.0	46	22	US-09-833-017-8	Sequence 8, Appli
3	46	100.0	46	22	US-09-833-017-9	Sequence 9, Appli
4	46	100.0	46	22	US-09-833-017-11	Sequence 11, Appl
5	46	100.0	46	22	US-09-833-017-12	Sequence 12, Appl
6	46	100.0	46	22	US-09-833-017-13	Sequence 13, Appl
7	41	89.1	46	22	US-09-833-017-7	Sequence 7, Appli

```
 8   27  58.7   43   22  US-09-833-017-10        US-09-833-017-10        Sequence  10, Appl
 9   21  45.7   21   24  US-09-833-017-14        US-09-833-017-14        Sequence  14, Appl
10    7  15.2   46   23  US-10-029-386-31630     US-10-029-386-31630     Sequence  31630, A
11    7  15.2  118   23  US-09-902-540-12650     US-09-902-540-12650     Sequence  12650, A
12    7  15.2  183   21  US-09-708-427-66071     US-09-708-427-66071     Sequence  66071, A
13    7  15.2  229   21  US-09-708-427-66070     US-09-708-427-66070     Sequence  66070, A
14    7  15.2  235   21  US-09-708-427-66069     US-09-708-427-66069     Sequence  66069, A
15    7  15.2  352    1  PCT-US01-14827-12064    PCT-US01-14827-12064    Sequence  12064, A
16    7  15.2  353   21  US-09-708-427-83339     US-09-708-427-83339     Sequence  83339, A
17    7  15.2  393   26  US-60-167-217-1542      US-60-167-217-1542      Sequence  1542, Ap
18    7  15.2  393   26  US-60-173-464-1254      US-60-173-464-1254      Sequence  1254, Ap
19    7  15.2  397   23  US-09-902-540-16267     US-09-902-540-16267     Sequence  16267, A
20    7  15.2  427   21  US-09-708-427-83338     US-09-708-427-83338     Sequence  83338, A
21    7  15.2  430   21  US-09-708-427-83337     US-09-708-427-83337     Sequence  83337, A
22    7  15.2  447   17  US-09-328-352-7917      US-09-328-352-7917      Sequence  7917, Ap
23    7  15.2  457   16  US-09-252-991A-30850    US-09-252-991A-30850    Sequence  30850, A
24    7  15.2  558    -  US-07-862-987A-2        US-07-862-987A-2        Sequence  2, Appli
25    7  15.2  586   21  US-09-708-427-5170      US-09-708-427-5170      Sequence  5170, Ap
26    7  15.2  612   21  US-09-708-427-5169      US-09-708-427-5169      Sequence  5169, Ap
27    7  15.2  618   21  US-09-708-427-5168      US-09-708-427-5168      Sequence  5168, Ap
28    7  15.2  658    1  PCT-US01-0863-48340     PCT-US01-0863-48340     Sequence  48340, A
29    7  15.2  767   20  US-09-606-304-8         US-09-606-304-8         Sequence  8, Appli
30    7  15.2  767   20  US-09-614-150-1512      US-09-614-150-1512      Sequence  1512, Ap
31    7  15.2  767   26  US-60-191-637-1515      US-60-191-637-1515      Sequence  1515, Ap
32    7  15.2  767   26  US-60-191-681-1203      US-60-191-681-1203      Sequence  1203, Ap
33    7  15.2  929   20  US-09-614-150-39363     US-09-614-150-39363     Sequence  39363, A
34    7  15.2  929   26  US-60-173-464-29732     US-60-173-464-29732     Sequence  29732, A
35    7  15.2  929   26  US-60-191-637-38989     US-60-191-637-38989     Sequence  38989, A
36    7  15.2  929   26  US-60-191-68-30211      US-60-191-68-30211      Sequence  30211, A
37    7  15.2 1357   16  US-09-270-767-44971     US-09-270-767-44971     Sequence  44971, A
38    7  15.2 1394   16  US-09-248-796-19555     US-09-248-796-19555     Sequence  19555, A
39    7  15.2 1394   26  US-60-096-405-19555     US-60-096-405-19555     Sequence  19555, A
40    7  15.2 1767   26  US-60-230-445-1919      US-60-230-445-1919      Sequence  1919, Ap
41    7  15.2 1773   26  US-60-230-445-1433      US-60-230-445-1433      Sequence  1433, Ap
42    7  15.2 2771   20  US-09-614-150-18432     US-09-614-150-18432     Sequence  18432, A
43    7  15.2 2771   26  US-60-167-217-18522     US-60-167-217-18522     Sequence  18522, A
44    7  15.2 2771   26  US-60-173-464-15176     US-60-173-464-15176     Sequence  15176, A
45    7  15.2 2771   26  US-60-191-637-18488     US-60-191-637-18488     Sequence  18488, A
46    7  15.2 2771   26  US-60-191-681-14627     US-60-191-681-14627     Sequence  14627, A
47    7  15.2 576775 12  US-09-270-767-2         US-09-270-767-2         Sequence  2, Appli
48    7  15.2 576775 12  US-08-895-611D-2        US-08-895-611D-2        Sequence  2, Appli
49    7  15.2 576775 22  US-09-895-61D-2         US-09-895-61D-2         Sequence  2, Appli
50    6  13.0   19   19  US-09-549-676A-21       US-09-549-676A-21       Sequence  21, Appl
51    6  13.0   20   22  US-09-895-828-472       US-09-895-828-472       Sequence  472, App
52    6  13.0   20   22  US-09-895-828-473       US-09-895-828-473       Sequence  473, App
53    6  13.0   20   21  US-09-614-150-18541     US-09-614-150-18541     Sequence  18541,
54    6  13.0   23   21  US-09-724-059-118730    US-09-724-059-118730    Sequence  118730,
55    6  13.0   23   21  US-09-724-059-118910    US-09-724-059-118910    Sequence  118910,
56    6  13.0   23   21  US-09-724-059-119099    US-09-724-059-119099    Sequence  119099,
57    6  13.0   23   21  US-09-724-059-119288    US-09-724-059-119288    Sequence  119288,
58    6  13.0   23   21  US-09-724-059-119459    US-09-724-059-119459    Sequence  119459,
59    6  13.0   23   21  US-09-724-059-119666    US-09-724-059-119666    Sequence  119666,
60    6  13.0   23   21  US-09-724-059-119837    US-09-724-059-119837    Sequence  119837,
61    6  13.0   23   21  US-09-724-059-120017    US-09-724-059-120017    Sequence  120017,
62    6  13.0   23   21  US-09-724-059-120206    US-09-724-059-120206    Sequence  120206,
63    6  13.0   23   21  US-09-724-059-120428    US-09-724-059-120428    Sequence  120428,
64    6  13.0   23   21  US-09-724-059-120806    US-09-724-059-120806    Sequence  120806,
65    6  13.0   23   21  US-09-724-059-121166    US-09-724-059-121166    Sequence  121166,
66    6  13.0   23   21  US-09-724-059-121544    US-09-724-059-121544    Sequence  121544,
67    6  13.0   23   21  US-09-724-059-121922    US-09-724-059-121922    Sequence  121922,
68    6  13.0   23   21  US-09-724-059-122264    US-09-724-059-122264    Sequence  122264,
69    6  13.0   23   21  US-09-724-059-122678    US-09-724-059-122678    Sequence  122678,
70    6  13.0   23   21  US-09-724-059-123020    US-09-724-059-123020    Sequence  123020,
71    6  13.0   23   21  US-09-724-059-123380    US-09-724-059-123380    Sequence  123380,
72    6  13.0   23   21  US-09-724-059-123758    US-09-724-059-123758    Sequence  123758,
73    6  13.0   23   21  US-09-724-059-127134    US-09-724-059-127134    Sequence  127134,
74    6  13.0   23   21  US-09-724-059-127260    US-09-724-059-127260    Sequence  127260,
75    6  13.0   23   21  US-09-724-059-127380    US-09-724-059-127380    Sequence  127380,
76    6  13.0   23   21  US-09-724-059-127506    US-09-724-059-127506    Sequence  127506,
77    6  13.0   23   21  US-09-724-059-127632    US-09-724-059-127632    Sequence  127632,
78    6  13.0   23   21  US-09-724-059-127746    US-09-724-059-127746    Sequence  127746,
79    6  13.0   23   21  US-09-724-059-127884    US-09-724-059-127884    Sequence  127884,
80    6  13.0   23   21  US-09-724-059-127998    US-09-724-059-127998    Sequence  127998,
```

#			N	Application	Sequence
81	13.0	6	23	US-09-724-059-128118	Sequence 128118,
82	13.0	6	21	US-09-724-059-128244	Sequence 128244,
83	13.0	6	21	US-09-724-059-128392	Sequence 128392,
84	13.0	6	21	US-09-724-059-128644	Sequence 128644,
85	13.0	6	21	US-09-724-059-128884	Sequence 128884,
86	13.0	6	21	US-09-724-059-129136	Sequence 129136,
87	13.0	6	21	US-09-724-059-129388	Sequence 129388,
88	13.0	6	21	US-09-724-059-129616	Sequence 129616,
89	13.0	6	21	US-09-724-059-129892	Sequence 129892,
90	13.0	6	21	US-09-724-059-130120	Sequence 130120,
91	13.0	6	21	US-09-724-059-130360	Sequence 130360,
92	13.0	6	21	US-09-724-059-130612	Sequence 130612,
93	13.0	6	21	US-09-724-059-139327	Sequence 139327,
94	13.0	6	21	US-09-724-059-140587	Sequence 140587,
95	13.0	6	21	US-09-724-059-141787	Sequence 141787,
96	13.0	6	21	US-09-724-059-143047	Sequence 143047,
97	13.0	6	21	US-09-724-059-144307	Sequence 144307,
98	13.0	6	21	US-09-724-059-145447	Sequence 145447,
99	13.0	6	21	US-09-724-059-146827	Sequence 146827,
100	13.0	6	21	US-09-724-059-147967	Sequence 147967,
101	13.0	6	21	US-09-724-059-149167	Sequence 149167,
102	13.0	6	21	US-09-724-059-150427	Sequence 150427,
103	13.0	6	21	US-09-724-059-151907	Sequence 151907,
104	13.0	6	21	US-09-724-059-154427	Sequence 154427,
105	13.0	6	21	US-09-724-059-156827	Sequence 156827,
106	13.0	6	21	US-09-724-059-159347	Sequence 159347,
107	13.0	6	21	US-09-724-059-161867	Sequence 161867,
108	13.0	6	21	US-09-724-059-164147	Sequence 164147,
109	13.0	6	21	US-09-724-059-166907	Sequence 166907,
110	13.0	6	21	US-09-724-059-169187	Sequence 169187,
111	13.0	6	21	US-09-724-059-171587	Sequence 171587,
112	13.0	6	21	US-09-724-059-174107	Sequence 174107,
113	13.0	6	26	US-60-143-992-840	Sequence 840, App
114	13.0	6	1	PCT-US99-03939-164	Sequence 164, App
115	13.0	6	27	US-09-716-129-168	Sequence 168, App
116	13.0	6	1	PCT-US01-00663-27981	Sequence 27981, A
117	13.0	6	30	US-09-864-761-34974	Sequence 34974, A
118	13.0	6	30	US-60-236-359-16494	Sequence 16494, A
119	13.0	6	1	PCT-US01-01354-18951	Sequence 18951, A
120	13.0	6	34	US-09-764-905-18951	Sequence 18951, A
121	13.0	6	34	US-10-092-399-18951	Sequence 18951, A
122	13.0	6	26	US-60-163-233-2268	Sequence 2268, Ap
123	13.0	6	3	US-07-906-308-10	Sequence 10, Appl
124	13.0	6	3	US-07-906-320-33	Sequence 33, Appl
125	13.0	6	40	US-60-162-243-477	Sequence 477, App
126	13.0	6	40	US-60-169-840-8280	Sequence 8280, Ap
127	13.0	6	42	US-60-195-300-15	Sequence 15, Appl
128	13.0	6	43	US-09-195-300-12	Sequence 12, Appl
129	13.0	6	45	US-60-182-302-608	Sequence 608, App
130	13.0	6	45	US-60-198-954-608	Sequence 608, App
131	13.0	6	48	US-08-756-693-5	Sequence 5, Appli
132	13.0	6	11	US-08-756-693A-5	Sequence 5, Appli
133	13.0	6	7	US-08-330-433-9	Sequence 9, Appli
134	13.0	6	51	US-08-330-433-11	Sequence 11, Appl
135	13.0	6	53	PCT-US99-22853B-2932	Sequence 2932, Ap
136	13.0	6	54	PCT-US00-26524B-7767	Sequence 7767, Ap
137	13.0	6	1	PCT-US00-18198-39	Sequence 39, Appl
138	13.0	6	59	US-09-345-464-39	Sequence 39, Appl
139	13.0	6	17	US-09-606-317-39	Sequence 39, Appl
140	13.0	6	20	US-09-834-366-23839	Sequence 23839, A
141	13.0	6	22	US-09-197-873-23839	Sequence 23839, A
142	13.0	6	59	US-60-197-712-2630	Sequence 2630, Ap
143	13.0	6	60	US-60-196-712-2630	Sequence 2630, Ap
144	13.0	6	7	US-08-330-433-10	Sequence 10, Appl
145	13.0	6	21	US-09-758-472-5612	Sequence 5612, Ap
146	13.0	6	26	US-60-178-310-333	Sequence 333, App
147	13.0	6	26	US-60-187-185-481	Sequence 481, App
148	13.0	6	16	US-09-252-691-10974	Sequence 10974, A
149	13.0	6	16	US-09-252-691C-10974	Sequence 10974, A
150	13.0	6	1	PCT-US01-06631-35062	Sequence 35062, A
151	13.0	6	1	PCT-US99-22853B-2931	Sequence 2931, Ap
152	13.0	6	7	US-08-330-433-12	Sequence 12, Appl
153	13.0	6	20	US-60-673-195A-264	Sequence 264, App
			26	US-60-177-571-4814	Sequence 4814, A

#			L	N	Application	Sequence
154	13.0	6	67	26	US-60-177-571-4860	Sequence 4860, Ap
155	13.0	6	69	1	PCT-US01-00663-31238	Sequence 31238, A
156	13.0	6	69	1	PCT-US01-14827-13422	Sequence 13422, A
157	13.0	6	69	22	US-09-864-761-38134	Sequence 38134, A
158	13.0	6	69	26	US-60-236-359-20765	Sequence 20765, A
159	13.0	6	70	1	PCT-US01-01324-2296	Sequence 2296, App
160	13.0	6	70	1	PCT-US01-0135-230	Sequence 230, App
161	13.0	6	70	21	US-09-764-887-230	Sequence 230, App
162	13.0	6	70	24	US-10-073-961-230	Sequence 230, App
163	13.0	6	70	24	US-10-079-979-2296	Sequence 2296, Ap
164	13.0	6	70	26	US-60-146-055-606	Sequence 606, App
165	13.0	6	71	25	US-60-186-661-437	Sequence 437, App
166	13.0	6	72	15	US-60-134-000-4826	Sequence 4826, Ap
167	13.0	6	74	21	US-09-739-449-10267	Sequence 10267, A
168	13.0	6	74	22	US-09-803-110-10267	Sequence 10267, A
169	13.0	6	75	26	US-60-169-840-4921	Sequence 4921, Ap
170	13.0	6	76	18	US-09-489-039A-7560	Sequence 7560, Ap
171	13.0	6	78	1	PCT-US01-0135-16970	Sequence 16970, A
172	13.0	6	78	21	US-09-764-905-16970	Sequence 16970, A
173	13.0	6	78	24	US-10-092-399-16970	Sequence 16970, A
174	13.0	6	79	1	PCT-US02-0510-964	Sequence 964, App
175	13.0	6	82	21	US-09-760-476-1727	Sequence 1727, Ap
176	13.0	6	83	1	PCT-US01-0133-5236	Sequence 5236, Ap
177	13.0	6	83	16	US-09-764-874-5236	Sequence 5236, Ap
178	13.0	6	85	16	US-09-270-767-41443	Sequence 41443, A
179	13.0	6	86	22	US-09-826-734-190	Sequence 190, App
180	13.0	6	86	26	US-60-233-940-132	Sequence 132, App
181	13.0	6	88	22	US-60-186-661-434	Sequence 434, App
182	13.0	6	88	26	US-60-192-737-1123	Sequence 1123, Ap
183	13.0	6	88	26	US-60-194-106-698	Sequence 698, App
184	13.0	6	89	26	US-60-186-661-440	Sequence 440, App
185	13.0	6	93	16	US-09-270-767-57065	Sequence 57065, A
186	13.0	6	93	16	US-09-270-849B-180996	Sequence 180996,
187	13.0	6	96	26	US-60-192-737-1113	Sequence 1113, Ap
188	13.0	6	96	26	US-60-194-106-688	Sequence 688, App
189	13.0	6	100	26	US-60-178-305-929	Sequence 929, App
190	13.0	6	100	26	US-60-178-305-931	Sequence 931, App
191	13.0	6	101	1	PCT-US00-05988-1709	Sequence 1709, Ap
192	13.0	6	101	23	US-09-925-300-1709	Sequence 1709, Ap
193	13.0	6	103	26	US-60-139-6-485	Sequence 485, A
194	13.0	6	105	1	PCT-US01-08631-44150	Sequence 44150, A
195	13.0	6	105	17	US-09-389-956-70	Sequence 70, Appl
196	13.0	6	105	20	US-09-675-784A-12517	Sequence 12517, A
197	13.0	6	110	13	US-08-905-135-467	Sequence 467, App
198	13.0	6	112	19	US-09-583-110-3690	Sequence 3690, Ap
199	13.0	6	114	15	US-09-150-816-52	Sequence 52, Appl
200	13.0	6	115	26	US-60-178-305-1127	Sequence 1127, Ap
201	13.0	6	116	15	US-60-107-443-4298	Sequence 4298, Ap
202	13.0	6	116	26	US-60-182-467-1968	Sequence 1968, Ap
203	13.0	6	119	16	US-09-270-767-44716	Sequence 44716, A
204	13.0	6	121	13	US-08-993-337-4996	Sequence 4996, Ap
205	13.0	6	121	17	US-09-330-337-321	Sequence 321, App
206	13.0	6	121	18	US-09-417-507-41433	Sequence 41433, A
207	13.0	6	122	26	US-60-182-567-612	Sequence 612, App
208	13.0	6	122	1	PCT-US99-03939-158	Sequence 158, App
209	13.0	6	122	13	US-08-993-002A-4997	Sequence 4997, Ap
210	13.0	6	122	16	US-09-248-796-24846	Sequence 24846, A
211	13.0	6	122	21	US-09-716-129-162	Sequence 162, App
212	13.0	6	123	23	US-09-902-540-11364	Sequence 11364, A
213	13.0	6	124	26	US-60-096-409-24846	Sequence 24846, A
214	13.0	6	124	26	US-60-178-305-1131	Sequence 1131, Ap
215	13.0	6	129	26	US-60-192-737-1114	Sequence 1114, Ap
216	13.0	6	130	26	US-60-194-106-689	Sequence 689, App
217	13.0	6	131	26	US-60-178-305-930	Sequence 930, App
218	13.0	6	132	13	US-08-993-002A-8527	Sequence 8527, Ap
219	13.0	6	132	13	US-08-993-002A-8526	Sequence 8526, Ap
220	13.0	6	132	8	PCT-US01-01347-346	Sequence 346, App
221	13.0	6	132	9	US-08-487-032A-790	Sequence 790, App
222	13.0	6	132	11	US-08-561-469A-790	Sequence 790, App
223	13.0	6	132	12	US-08-761-184-1029	Sequence 1029, Ap
224	13.0	6	132	12	US-08-821-931-1029	Sequence 1029, Ap
225	13.0	6	132	13	US-08-993-002A-7629	Sequence 7629, Ap
226	13.0	6	133	1	PCT-US01-14827-15643	Sequence 15643, A

#					Accession	Sequence
227	6	13.0	134	1	PCT-US01-14827-15658	Sequence 15658, A
228	6	13.0	135	26	US-60-182-567-614	Sequence 614, App
229	6	13.0	135	26	US-60-188-162-3521	Sequence 3521, Ap
230	6	13.0	138	21	US-09-733-089-14808	Sequence 14808, A
231	6	13.0	138	22	US-09-816-660-14808	Sequence 14808, A
232	6	13.0	139	17	US-09-358-322-11	Sequence 11, Appl
233	6	13.0	139	17	US-09-358-322A-11	Sequence 11, Appl
234	6	13.0	139	24	US-10-004-494A-11	Sequence 11, Appl
235	6	13.0	141	26	US-60-207-213-271	Sequence 271, App
236	6	13.0	141	15	US-09-198-452A-1236	Sequence 1236, Ap
237	6	13.0	147	1	PCT-US01-05363-7	Sequence 7, Appli
238	6	13.0	147	1	PCT-US01-05578-7	Sequence 7, Appli
239	6	13.0	147	21	US-09-708-427-14541	Sequence 14541, A
240	6	13.0	147	21	US-09-725-413A-4	Sequence 4, Appli
241	6	13.0	147	21	US-09-789-836-7	Sequence 7, Appli
242	6	13.0	148	15	US-09-134-001C-4679	Sequence 4679, Ap
243	6	13.0	148	16	US-09-270-767-38744	Sequence 38744, A
244	6	13.0	148	16	US-09-270-767-53961	Sequence 53961, A
245	6	13.0	148	16	US-09-270-849B-189782	Sequence 189782,
246	6	13.0	148	16	US-09-450-969-6155	Sequence 6155, Ap
247	6	13.0	148	18	US-09-450-969-7130	Sequence 7130, Ap
248	6	13.0	149	1	PCT-US01-08631-47633	Sequence 47633, A
249	6	13.0	150	1	PCT-US01-01341-1052	Sequence 1052, Ap
250	6	13.0	150	18	US-09-438-185-956	Sequence 956, App
251	6	13.0	150	18	US-09-438-185A-956	Sequence 956, App
252	6	13.0	150	21	US-09-764-864-1052	Sequence 1052, Ap
253	6	13.0	150	24	US-10-080-129-1052	Sequence 1052, Ap
254	6	13.0	150	26	US-60-141-863-678	Sequence 678, App
255	6	13.0	150	26	US-60-141-863A-678	Sequence 678, App
256	6	13.0	151	16	US-09-270-767-42527	Sequence 42527, A
257	6	13.0	154	26	US-60-170-374-3444	Sequence 3444, Ap
258	6	13.0	158	26	US-60-178-305-1125	Sequence 1125, Ap
259	6	13.0	160	16	US-09-270-767-36332	Sequence 36332, A
260	6	13.0	160	16	US-09-270-767-51549	Sequence 51549, A
261	6	13.0	160	16	US-09-270-849B-184782	Sequence 184782,
262	6	13.0	160	18	US-09-489-039A-8791	Sequence 8791, Ap
263	6	13.0	160	20	US-09-614-450-701	Sequence 701, App
264	6	13.0	160	21	US-09-708-427-60528	Sequence 60528, A
265	6	13.0	161	18	US-09-417-507-37306	Sequence 37306, A
266	6	13.0	162	26	US-60-192-737-1116	Sequence 1116, Ap
267	6	13.0	162	26	US-60-194-106-691	Sequence 691, App
268	6	13.0	163	11	US-08-792-511-17	Sequence 17, Appl
269	6	13.0	163	16	US-09-252-691-5947	Sequence 5947, Ap
270	6	13.0	163	16	US-09-252-691C-5947	Sequence 5947, Ap
271	6	13.0	166	16	US-09-270-767-33483	Sequence 33483, A
272	6	13.0	166	16	US-09-270-767-48700	Sequence 48700, A
273	6	13.0	166	16	US-09-270-849B-192226	Sequence 192226,
274	6	13.0	167	26	US-60-192-737-1117	Sequence 1117, Ap
275	6	13.0	167	26	US-60-194-106-692	Sequence 692, App
276	6	13.0	168	26	US-60-192-737-1121	Sequence 1121, Ap
277	6	13.0	168	26	US-60-194-106-696	Sequence 696, App
278	6	13.0	169	26	US-60-182-567-613	Sequence 613, App
279	6	13.0	171	26	US-60-192-737-1122	Sequence 1122, Ap
280	6	13.0	171	26	US-60-194-106-697	Sequence 697, App
281	6	13.0	172	1	PCT-US01-01341-1315	Sequence 1315, Ap
282	6	13.0	172	21	US-09-764-864-1315	Sequence 1315, Ap
283	6	13.0	172	24	US-10-080-129-1315	Sequence 1315, Ap
284	6	13.0	174	10	US-08-625-811-1685	Sequence 1685, Ap
285	6	13.0	174	13	US-09-893-002A-5416	Sequence 5416, Ap
286	6	13.0	174	20	US-09-614-450-700	Sequence 700, App
287	6	13.0	174	21	US-09-708-427-60527	Sequence 60527, A
288	6	13.0	175	26	US-60-186-661-439	Sequence 439, App
289	6	13.0	178	26	US-60-178-305-1126	Sequence 1126, Ap
290	6	13.0	180	11	US-08-756-693-2	Sequence 2, Appli
291	6	13.0	180	11	US-08-756-693A-2	Sequence 2, Appli
292	6	13.0	183	26	US-60-192-737-1115	Sequence 1115, Ap
293	6	13.0	183	26	US-60-194-106-690	Sequence 690, App
294	6	13.0	184	26	US-60-142-896-1235	Sequence 1235, Ap
295	6	13.0	185	26	US-60-161-932-2238	Sequence 2238, Ap
296	6	13.0	186	11	US-08-756-693-1	Sequence 1, Appli
297	6	13.0	186	11	US-08-756-693-9	Sequence 9, Appli
298	6	13.0	186	11	US-08-756-693A-9	Sequence 9, Appli
299	6	13.0	186	21	US-09-708-427-14540	Sequence 14540, A
300	6	13.0	186	26	US-60-150-584-1011	Sequence 1011, Ap
301	6	13.0	186	26	US-60-192-737-1118	Sequence 1118, Ap
302	6	13.0	186	26	US-60-194-106-693	Sequence 693, App
303	6	13.0	191	26	US-60-195-136-825	Sequence 825, App
304	6	13.0	195	26	US-60-213-845-344	Sequence 344, App
305	6	13.0	197	16	US-09-252-991A-28165	Sequence 28165, A
306	6	13.0	197	26	US-60-178-305-1128	Sequence 1128, Ap
307	6	13.0	198	16	US-09-270-767-61873	Sequence 61873, A
308	6	13.0	198	16	US-09-270-849B-193908	Sequence 193908,
309	6	13.0	198	21	US-09-708-427-21930	Sequence 21930, A
310	6	13.0	198	21	US-09-708-427-57479	Sequence 57479, A
311	6	13.0	201	26	US-60-316-362-3153	Sequence 3153, Ap
312	6	13.0	203	19	US-09-583-110-3736	Sequence 3736, Ap
313	6	13.0	205	26	US-60-167-217-12329	Sequence 12329, A
314	6	13.0	208	21	US-09-760-475-1934	Sequence 1934, Ap
315	6	13.0	209	22	US-09-760-475-3224	Sequence 3224, Ap
316	6	13.0	211	26	US-09-870-759-39	Sequence 39, Appl
317	6	13.0	212	26	US-60-341-26-3037	Sequence 3037, Ap
318	6	13.0	214	15	PCT-US01-08631-55189	Sequence 55189, A
319	6	13.0	214	26	US-09-107-433-3882	Sequence 3882, Ap
320	6	13.0	214	26	US-60-192-737-1120	Sequence 1120, Ap
321	6	13.0	216	26	US-60-194-106-695	Sequence 695, App
322	6	13.0	216	26	US-60-192-737-1119	Sequence 1119, Ap
323	6	13.0	219	26	US-60-194-106-694	Sequence 694, App
324	6	13.0	219	26	US-60-178-305-1130	Sequence 1130, Ap
325	6	13.0	220	1	PCT-US98-0637-546	Sequence 546, App
326	6	13.0	220	12	US-08-881-227-176	Sequence 176, App
327	6	13.0	220	13	US-08-993-002A-5417	Sequence 5417, Ap
328	6	13.0	221	26	US-60-147-394-1698	Sequence 1698, Ap
329	6	13.0	221	26	US-60-147-189-977	Sequence 977, App
330	6	13.0	222	16	US-60-178-305-1129	Sequence 1129, Ap
331	6	13.0	222	16	US-09-201-228A-1005	Sequence 1005, Ap
332	6	13.0	222	26	US-09-140-956-1453	Sequence 1453, Ap
333	6	13.0	223	21	US-09-708-427-21929	Sequence 21929, A
334	6	13.0	223	21	US-09-708-427-57478	Sequence 57478, A
335	6	13.0	224	20	US-09-614-150-12297	Sequence 12297, A
336	6	13.0	224	26	US-60-185-360-244	Sequence 244, App
337	6	13.0	224	26	US-60-186-661-435	Sequence 435, App
338	6	13.0	224	26	US-60-186-661-438	Sequence 438, App
339	6	13.0	224	26	US-60-191-637-12332	Sequence 12332, A
340	6	13.0	225	11	US-08-756-693-8	Sequence 8, Appli
341	6	13.0	225	11	US-08-756-693A-8	Sequence 8, Appli
342	6	13.0	225	26	US-60-186-661-436	Sequence 436, App
343	6	13.0	225	26	US-60-243-468-1247	Sequence 1247, Ap
344	6	13.0	226	15	US-09-107-532-3948	Sequence 3948, Ap
345	6	13.0	226	15	US-09-107-532A-3948	Sequence 3948, Ap
346	6	13.0	228	26	US-09-322-109-27251	Sequence 27251, A
347	6	13.0	231	21	US-09-733-089-1021	Sequence 1021, Ap
348	6	13.0	231	22	US-09-816-660-1021	Sequence 1021, Ap
349	6	13.0	233	21	US-09-708-427-21928	Sequence 21928, A
350	6	13.0	233	23	US-60-312-544-6615	Sequence 6615, App
351	6	13.0	236	23	US-09-902-540-15029	Sequence 15029, A
352	6	13.0	238	26	US-60-185-360-246	Sequence 246, App
353	6	13.0	242	26	US-60-128-476-2875	Sequence 2875, Ap
354	6	13.0	243	15	US-09-198-452A-284	Sequence 284, App
355	6	13.0	243	21	US-09-708-427-57477	Sequence 57477, A
356	6	13.0	243	21	US-09-760-475-2085	Sequence 2085, Ap
357	6	13.0	245	18	US-09-438-185-273	Sequence 273, App
358	6	13.0	245	18	US-09-438-185A-273	Sequence 273, App
359	6	13.0	247	19	US-09-543-681A-5522	Sequence 5522, Ap
360	6	13.0	248	26	US-60-217-7825	Sequence 7825, Ap
361	6	13.0	248	26	US-60-173-464-6250	Sequence 6250, Ap
362	6	13.0	248	26	US-60-285-697-34	Sequence 34, Appl
363	6	13.0	248	26	US-60-295-890-34	Sequence 34, Appl
364	6	13.0	252	26	PCT-US01-08631-44716	Sequence 44716, A
365	6	13.0	252	21	PCT-US01-18569-2358	Sequence 2358, Ap
366	6	13.0	252	21	US-09-760-475-2092	Sequence 2092, Ap
367	6	13.0	254	1	PCT-US01-08631-43146	Sequence 43146, A
368	6	13.0	254	22	PCT-US01-08631-46154	Sequence 46154, A
369	6	13.0	255	26	US-09-895-828-454	Sequence 454, App
370	6	13.0	255	26	US-60-140-956-1355	Sequence 1355, Ap
371	6	13.0	255	26	US-60-324-631-1507	Sequence 1507, Ap
372	6	13.0	261	26	US-60-167-217-13019	Sequence 13019, A

#					Reference	Sequence
373	6	13.0	261	26	US-60-173-464-10586	Sequence 10586, A
374	6	13.0	261	26	US-60-191-637-13013	Sequence 13013, A
375	6	13.0	261	26	US-60-191-681-10235	Sequence 10235, A
376	6	13.0	262	22	US-09-895-828-457	Sequence 457, App
377	6	13.0	265	26	US-09-708-427-30798	Sequence 30798, A
378	6	13.0	266	16	US-09-252-991A-19949	Sequence 19949, A
379	6	13.0	268	16	US-09-248-796-17195	Sequence 17195, A
380	6	13.0	268	26	US-60-191-637-17195	Sequence 17195, A
381	6	13.0	268	26	US-60-191-637-25193	Sequence 25193, A
382	6	13.0	268	26	US-60-191-681-19832	Sequence 19832, A
383	6	13.0	271	26	US-60-739-449-8327	Sequence 8327, App
384	6	13.0	271	22	US-09-803-110-8327	Sequence 8327, App
385	6	13.0	272	26	US-60-185-360-245	Sequence 245, App
386	6	13.0	273	11	US-08-761-184-1339	Sequence 1339, Ap
387	6	13.0	273	11	US-08-761-184-1620	Sequence 1620, Ap
388	6	13.0	273	12	US-08-821-931-1339	Sequence 1339, Ap
389	6	13.0	273	12	US-08-821-931-1620	Sequence 1620, Ap
390	6	13.0	273	13	US-08-993-002A-7630	Sequence 7630, Ap
391	6	13.0	273	13	US-08-993-002A-7631	Sequence 7631, Ap
392	6	13.0	276	20	US-09-614-150-35646	Sequence 35646, A
393	6	13.0	277	26	US-60-191-637-35257	Sequence 35257, A
394	6	13.0	277	1	PCT-US00-35017A-1060	Sequence 1060, Ap
395	6	13.0	277	1	PCT-US97-21976-187	Sequence 187, App
396	6	13.0	277	17	US-09-308-967-187	Sequence 187, App
397	6	13.0	277	17	US-09-308-967B-187	Sequence 187, App
398	6	13.0	279	1	PCT-US01-01334-5024	Sequence 5024, Ap
399	6	13.0	279	1	PCT-US01-01341-1477	Sequence 1477, Ap
400	6	13.0	279	21	US-09-764-864-1477	Sequence 1477, Ap
401	6	13.0	279	21	US-09-764-874-5024	Sequence 5024, Ap
402	6	13.0	279	24	US-10-080-129-1477	Sequence 1477, Ap
403	6	13.0	280	18	US-09-417-507-41171	Sequence 41171, A
404	6	13.0	280	20	US-09-643-192-141	Sequence 141, App
405	6	13.0	280	20	US-09-643-192A-141	Sequence 141, App
406	6	13.0	282	8	US-08-448-729-56	Sequence 56, Appl
407	6	13.0	282	8	US-08-449-182-56	Sequence 56, Appl
408	6	13.0	282	8	US-08-449-649-56	Sequence 56, Appl
409	6	13.0	282	13	US-08-993-002A-4789	Sequence 4789, Ap
410	6	13.0	282	14	US-09-005-243-56	Sequence 56, Appl
411	6	13.0	282	14	US-09-005-858-56	Sequence 56, Appl
412	6	13.0	282	14	US-09-005-893-56	Sequence 56, Appl
413	6	13.0	282	14	US-09-005-907-56	Sequence 56, Appl
414	6	13.0	282	14	US-09-005-917-56	Sequence 56, Appl
415	6	13.0	282	16	US-09-083-838-56	Sequence 56, Appl
416	6	13.0	282	16	US-09-224-683-56	Sequence 56, Appl
417	6	13.0	282	20	US-09-635-249-56	Sequence 56, Appl
418	6	13.0	282	20	US-09-635-251-56	Sequence 56, Appl
419	6	13.0	282	20	US-09-643-658A-56	Sequence 56, Appl
420	6	13.0	282	20	US-09-643-659B-56	Sequence 56, Appl
421	6	13.0	282	21	US-09-758-444-572	Sequence 572, App
422	6	13.0	283	11	US-08-761-184-1107	Sequence 1107, Ap
423	6	13.0	283	12	US-08-821-931-1107	Sequence 1107, Ap
424	6	13.0	283	13	US-08-993-002A-4788	Sequence 4788, Ap
425	6	13.0	284	16	US-09-270-767-58734	Sequence 58734, A
426	6	13.0	286	16	US-09-270-849B-185550	Sequence 185550,
427	6	13.0	288	1	PCT-US01-08656-10098	Sequence 10098,
428	6	13.0	288	24	US-10-015-127-13977	Sequence 13977, A
429	6	13.0	289	1	PCT-US01-08631-53361	Sequence 53361, A
430	6	13.0	289	18	US-09-417-507-26016	Sequence 26016, A
431	6	13.0	289	26	US-09-614-150-9753	Sequence 9753, Ap
432	6	13.0	289	26	US-60-167-217-9791	Sequence 9791, Ap
433	6	13.0	289	26	US-60-173-464-7897	Sequence 7897, Ap
434	6	13.0	289	26	US-60-191-637-9785	Sequence 9785, Ap
435	6	13.0	289	26	US-60-191-681-7631	Sequence 7631, Ap
436	6	13.0	292	14	US-09-051-755-10	Sequence 10, Appl
437	6	13.0	294	20	US-09-902-540-12714	Sequence 12714, A
438	6	13.0	295	21	US-09-739-449-9863	Sequence 9863, Ap
439	6	13.0	295	22	US-09-803-110-9863	Sequence 9863, Ap
440	6	13.0	296	1	PCT-US01-05363-9	Sequence 9, Appli
441	6	13.0	296	1	PCT-US01-05578-9	Sequence 9, Appli
442	6	13.0	296	16	US-09-270-767-44827	Sequence 44827, A
443	6	13.0	296	21	US-09-789-836-9	Sequence 9, Appli
444	6	13.0	297	1	PCT-US01-05363-18	Sequence 18, Appl
445	6	13.0	297	1	PCT-US01-05578-11	Sequence 11, Appl

#					Reference	Sequence
446	6	13.0	21	297	US-09-739-449-11006	Sequence 11006, A
447	6	13.0	21	297	US-09-789-836-11	Sequence 11, Appl
448	6	13.0	22	297	US-09-803-110-11006	Sequence 11006, A
449	6	13.0	26	299	US-60-207-213-317	Sequence 317, App
450	6	13.0	16	303	US-09-270-767-42501	Sequence 42501, A
451	6	13.0	16	305	US-09-270-767-38800	Sequence 38800, A
452	6	13.0	16	305	US-09-270-767-54017	Sequence 54017, A
453	6	13.0	26	305	US-09-270-849B-189913	Sequence 189913,
454	6	13.0	26	305	US-60-208-318-11	Sequence 11, Appl
455	6	13.0	26	305	US-60-208-965-172	Sequence 172, App
456	6	13.0	26	305	US-60-248-505-837	Sequence 837, App
457	6	13.0	16	306	US-09-270-767-41820	Sequence 41820, A
458	6	13.0	26	307	US-09-324-109-30036	Sequence 30036, A
459	6	13.0	18	308	US-09-489-039A-7359	Sequence 7359, Ap
460	6	13.0	1	312	PCT-US01-08631-38763	Sequence 38763, A
461	6	13.0	22	312	US-09-897-516-6721	Sequence 6721, Ap
462	6	13.0	26	312	US-60-815-161-6721	Sequence 6721, Ap
463	6	13.0	26	312	US-60-324-109-22174	Sequence 22174, A
464	6	13.0	26	313	US-60-315-676-12	Sequence 12, Appl
465	6	13.0	20	317	US-09-614-150-35325	Sequence 35325, A
466	6	13.0	26	317	US-60-173-464-26963	Sequence 26963, A
467	6	13.0	26	317	US-60-191-637-34930	Sequence 34930, A
468	6	13.0	26	317	US-60-191-681-27412	Sequence 27412, A
469	6	13.0	1	321	PCT-US02-03987-5771	Sequence 5771, Ap
470	6	13.0	1	321	PCT-US02-03987-12694	Sequence 12694, A
471	6	13.0	1	321	PCT-US02-03987-12721	Sequence 12721, A
472	6	13.0	22	321	US-09-815-242-5771	Sequence 5771, Ap
473	6	13.0	22	321	US-09-815-242-12694	Sequence 12694, A
474	6	13.0	22	321	US-09-815-242-12721	Sequence 12721, A
475	6	13.0	23	321	US-09-924-159-163	Sequence 163, App
476	6	13.0	24	321	US-10-072-851-5771	Sequence 5771, Ap
477	6	13.0	24	321	US-10-072-851-12694	Sequence 12694, A
478	6	13.0	24	321	US-10-072-851-12721	Sequence 12721, A
479	6	13.0	26	321	US-60-242-578-1031	Sequence 1031, Ap
480	6	13.0	26	321	US-60-253-625-2375	Sequence 2375, Ap
481	6	13.0	1	322	PCT-US01-05363-6	Sequence 6, Appli
482	6	13.0	1	322	PCT-US01-05578-6	Sequence 6, Appli
483	6	13.0	21	322	US-09-789-836-6	Sequence 6, Appli
484	6	13.0	22	322	US-09-823-038A-33	Sequence 33, Appl
485	6	13.0	18	323	US-09-438-185-443	Sequence 443, App
486	6	13.0	18	323	US-09-438-185A-443	Sequence 443, App
487	6	13.0	26	324	US-60-315-676-9	Sequence 9, Appli
488	6	13.0	1	329	PCT-US01-08631-41597	Sequence 41597, A
489	6	13.0	1	329	PCT-US02-03987-11142	Sequence 11142, A
490	6	13.0	22	329	US-09-815-242-11142	Sequence 11142, A
491	6	13.0	24	329	US-10-072-851-11142	Sequence 11142, A
492	6	13.0	1	331	PCT-US01-08631-49770	Sequence 49770, A
493	6	13.0	19	331	US-09-500-299-14	Sequence 14, Appl
494	6	13.0	20	332	US-09-611-529-6398	Sequence 6398, Ap
495	6	13.0	26	333	US-60-185-361-840	Sequence 840, App
496	6	13.0	16	334	US-09-248-796-20300	Sequence 20300, A
497	6	13.0	26	334	US-60-096-409-20300	Sequence 20300, A
498	6	13.0	24	336	US-10-000-897-69	Sequence 69, Appl
499	6	13.0	21	338	US-09-739-449-10690	Sequence 10690, A
500	6	13.0	22	338	US-09-803-110-10690	Sequence 10690, A
501	6	13.0	1	339	PCT-US01-08631-57218	Sequence 57218, A
502	6	13.0	26	341	US-60-185-361-776	Sequence 776, App
503	6	13.0	11	342	US-08-761-184-1459	Sequence 1459, Ap
504	6	13.0	12	342	US-08-821-931-1459	Sequence 1459, Ap
505	6	13.0	13	342	US-08-993-002A-4790	Sequence 4790, Ap
506	6	13.0	13	342	US-08-993-002A-7045	Sequence 7045, Ap
507	6	13.0	20	342	US-09-614-150-33645	Sequence 33645, A
508	6	13.0	20	342	US-09-815-108-20	Sequence 20, Appl
509	6	13.0	26	342	US-60-173-464-22113	Sequence 22113, A
510	6	13.0	26	342	US-60-191-637-33223	Sequence 33223, A
511	6	13.0	26	342	US-60-191-681-26158	Sequence 26158, A
512	6	13.0	1	344	PCT-US01-08631-33558	Sequence 33558, A
513	6	13.0	21	344	US-08-712-563-165	Sequence 165, App
514	6	13.0	23	344	US-09-902-540-11424	Sequence 11424, A
515	6	13.0	1	345	PCT-US01-05363-5	Sequence 5, Appli
516	6	13.0	1	345	PCT-US01-05578-5	Sequence 5, Appli
517	6	13.0	1	345	PCT-US01-10661-10	Sequence 10, Appl
518	6	13.0	1	345	PCT-US02-02781-183	Sequence 183, App

Index			Len	N	Reference	Sequence
519	13.0	6	345	21	US-09-789-836-5	Sequence 5, Appli
520	13.0	6	345	24	US-10-060-036-183	Sequence 183, App
521	13.0	6	350	1	PCT-US00-18198-103	Sequence 103, App
522	13.0	6	350	1	PCT-US02-03987-15264	Sequence 15264, A
523	13.0	6	350	17	US-09-345-464-103	Sequence 103, App
524	13.0	6	350	20	US-09-606-317-103	Sequence 103, App
525	13.0	6	350	24	US-10-032-585-7320	Sequence 7320, Ap
526	13.0	6	350	24	US-10-072-851-15264	Sequence 15264, A
527	13.0	6	350	26	US-60-189-998-3	Sequence 3, Appli
528	13.0	6	350	26	US-60-259-128-4741	Sequence 4741, Ap
529	13.0	6	350	26	US-60-314-050-7320	Sequence 7320, Ap
530	13.0	6	351	8	US-08-487-032A-640	Sequence 640, App
531	13.0	6	351	9	US-08-561-469A-640	Sequence 640, App
532	13.0	6	351	11	US-08-761-184-983	Sequence 983, App
533	13.0	6	351	12	US-08-821-931-983	Sequence 983, App
534	13.0	6	351	13	US-08-993-002A-4795	Sequence 4795, Ap
535	13.0	6	352	20	US-09-699-146-244	Sequence 244, App
536	13.0	6	354	11	US-08-761-184-1560	Sequence 1560, Ap
537	13.0	6	354	12	US-08-821-931-1560	Sequence 1560, Ap
538	13.0	6	354	13	US-08-993-002A-4796	Sequence 4796, Ap
539	13.0	6	355	11	US-08-761-184-1161	Sequence 1161, Ap
540	13.0	6	355	12	US-08-821-931-1161	Sequence 1161, Ap
541	13.0	6	355	13	US-08-993-002A-4797	Sequence 4797, Ap
542	13.0	6	355	18	US-09-417-507-28013	Sequence 28013, A
543	13.0	6	356	18	US-09-488-725A-2917	Sequence 2917, Ap
544	13.0	6	357	16	US-09-252-991A-25036	Sequence 25036, A
545	13.0	6	358	23	US-09-933-480-3	Sequence 3, Appli
546	13.0	6	359	20	US-09-614-150-39777	Sequence 39777, A
547	13.0	6	359	26	US-09-146-315-913	Sequence 913, App
548	13.0	6	359	26	US-60-173-464-29633	Sequence 29633, A
549	13.0	6	359	26	US-60-173-464-29936	Sequence 29936, A
550	13.0	6	359	26	US-60-191-637-39409	Sequence 39409, A
551	13.0	6	359	26	US-60-191-681-30457	Sequence 30457, A
552	13.0	6	360	1	PCT-US01-08631-45056	Sequence 45056, A
553	13.0	6	360	15	US-09-198-452A-342	Sequence 342, App
554	13.0	6	360	17	US-09-376-770-6	Sequence 6, Appli
555	13.0	6	360	18	US-09-488-725A-6489	Sequence 6489, Ap
556	13.0	6	360	20	US-09-614-150-39240	Sequence 39240, A
557	13.0	6	360	26	US-60-191-637-38866	Sequence 38866, A
558	13.0	6	360	26	US-60-191-681-30124	Sequence 30124, A
559	13.0	6	365	1	PCT-US00-35017A-1026	Sequence 1026, Ap
560	13.0	6	366	18	US-09-438-185-324	Sequence 324, App
561	13.0	6	366	18	US-09-438-185A-324	Sequence 324, App
562	13.0	6	367	15	US-09-134-001C-3678	Sequence 3678, Ap
563	13.0	6	368	1	PCT-US00-30036A-108	Sequence 108, App
564	13.0	6	368	1	PCT-US01-11988-707	Sequence 707, App
565	13.0	6	368	1	PCT-US02-08123-1751	Sequence 1751, Ap
566	13.0	6	368	1	PCT-US02-08277-1146	Sequence 1146, Ap
567	13.0	6	368	1	PCT-US02-08277-1683	Sequence 1683, Ap
568	13.0	6	368	22	US-09-833-245-707	Sequence 707, App
569	13.0	6	370	15	US-09-107-433-3702	Sequence 3702, Ap
570	13.0	6	370	19	US-09-583-110-3405	Sequence 3405, Ap
571	13.0	6	371	12	US-08-827-356-3712	Sequence 3712, Ap
572	13.0	6	371	18	US-09-488-725A-5360	Sequence 5360, Ap
573	13.0	6	373	20	US-09-611-529-4501	Sequence 4501, Ap
574	13.0	6	373	22	US-09-823-038A-60	Sequence 60, Appl
575	13.0	6	375	21	US-09-708-427-10518	Sequence 10518, A
576	13.0	6	377	1	PCT-US01-26213-2	Sequence 2, Appli
577	13.0	6	377	13	US-09-988-671A-31	Sequence 31, Appl
578	13.0	6	377	16	US-09-270-767-46304	Sequence 46304, A
579	13.0	6	377	18	US-09-488-725A-1978	Sequence 1978, Ap
580	13.0	6	377	20	US-09-611-526-2593	Sequence 2593, Ap
581	13.0	6	377	23	US-09-680-985-2	Sequence 2, Appli
582	13.0	6	377	23	US-09-935-038-2	Sequence 2, Appli
583	13.0	6	377	23	US-09-935-390A-31	Sequence 31, Appl
584	13.0	6	377	24	US-10-094-474-2	Sequence 2, Appli
585	13.0	6	377	26	US-60-213-845-443	Sequence 443, App
586	13.0	6	378	22	US-09-833-790-256	Sequence 256, App
587	13.0	6	385	18	US-09-488-725A-5550	Sequence 5550, Ap
588	13.0	6	385	19	US-09-570-581A-885	Sequence 885, App
589	13.0	6	388	26	US-60-315-676-2034	Sequence 2034, Ap
590	13.0	6	390	26	US-60-221-680-5	Sequence 5, Appli
591	13.0	6	392	20	US-09-614-150-1503	Sequence 1503, Ap

Index			Len	N	Reference	Sequence
592	13.0	6	392	20	US-09-619-049-108	Sequence 108, App
593	13.0	6	392	26	US-60-167-217-1530	Sequence 1530, Ap
594	13.0	6	392	26	US-60-171-627-144	Sequence 144, App
595	13.0	6	392	26	US-60-173-464-1242	Sequence 1242, Ap
596	13.0	6	392	26	US-60-191-637-1506	Sequence 1506, Ap
597	13.0	6	392	26	US-60-191-681-1194	Sequence 1194, Ap
598	13.0	6	395	22	US-09-897-516-5816	Sequence 5816, Ap
599	13.0	6	395	26	US-60-215-081-5816	Sequence 5816, Ap
600	13.0	6	396	18	US-60-470-154A-49	Sequence 49, Appli
601	13.0	6	397	26	US-60-179-903-2	Sequence 2, Appli
602	13.0	6	398	1	PCT-US01-08631-41200	Sequence 41200, A
603	13.0	6	404	1	PCT-US00-26524B-6307	Sequence 6307, Ap
604	13.0	6	404	20	US-09-614-150-21411	Sequence 21411, A
605	13.0	6	404	26	US-60-167-217-21620	Sequence 21620, A
606	13.0	6	404	26	US-60-173-464-17585	Sequence 17585, A
607	13.0	6	404	26	US-60-191-637-21473	Sequence 21473, A
608	13.0	6	404	26	US-60-191-681-16925	Sequence 16925, A
609	13.0	6	406	1	PCT-US01-14827-11742	Sequence 11742, A
610	13.0	6	410	1	PCT-US00-30036A-84	Sequence 84, Appl
611	13.0	6	410	1	PCT-US01-11988-710	Sequence 710, App
612	13.0	6	410	1	PCT-US02-08123-1365	Sequence 1365, Ap
613	13.0	6	410	1	PCT-US02-08277-890	Sequence 890, App
614	13.0	6	410	1	PCT-US02-08278-1312	Sequence 1312, Ap
615	13.0	6	410	22	US-09-833-245-710	Sequence 710, App
616	13.0	6	412	9	US-08-553-125-12	Sequence 12, Appl
617	13.0	6	412	9	US-08-553-125A-12	Sequence 12, Appl
618	13.0	6	412	11	US-08-798-095-12	Sequence 12, Appl
619	13.0	6	412	11	US-08-798-095A-12	Sequence 12, Appl
620	13.0	6	412	11	US-08-798-096-12	Sequence 12, Appl
621	13.0	6	412	13	US-08-915-095-12	Sequence 12, Appl
622	13.0	6	412	13	US-08-915-095A-12	Sequence 12, Appl
623	13.0	6	412	13	US-08-948-538A-41	Sequence 41, Appl
624	13.0	6	412	16	US-09-215-450-25	Sequence 25, Appl
625	13.0	6	412	16	US-09-219-441-12	Sequence 12, Appl
626	13.0	6	412	16	US-09-219-441A-12	Sequence 12, Appl
627	13.0	6	412	17	US-09-380-159-6	Sequence 6, Appli
628	13.0	6	412	22	US-09-882-636-6	Sequence 6, Appli
629	13.0	6	412	23	US-09-953-956-412	Sequence 12, Appl
630	13.0	6	413	22	US-09-897-516-6849	Sequence 6849, Ap
631	13.0	6	413	26	US-60-173-464-3850	Sequence 3850, Ap
632	13.0	6	413	26	US-60-173-464-3176	Sequence 3176, Ap
633	13.0	6	413	26	US-60-191-637-6849	Sequence 6849, Ap
634	13.0	6	414	26	US-60-215-161-6849	Sequence 6849, Ap
635	13.0	6	414	26	US-60-245-228-383	Sequence 383, App
636	13.0	6	418	26	US-60-213-177-637	Sequence 637, App
637	13.0	6	418	26	US-60-185-361-534	Sequence 534, App
638	13.0	6	422	21	US-60-733-089-15567	Sequence 15567, A
639	13.0	6	422	22	US-09-816-660-15567	Sequence 15567, A
640	13.0	6	423	20	US-09-614-150-14088	Sequence 14088, A
641	13.0	6	423	26	US-60-191-637-14129	Sequence 14129, A
642	13.0	6	435	16	US-09-252-691-9032	Sequence 9032, Ap
643	13.0	6	435	16	US-09-252-691C-9032	Sequence 9032, Ap
644	13.0	6	435	26	US-60-167-216-74	Sequence 74, Appl
645	13.0	6	435	26	US-60-173-464-2285	Sequence 2285, Ap
646	13.0	6	438	1	PCT-US01-4782-2	Sequence 2, Appli
647	13.0	6	438	24	US-10-001-848-2	Sequence 2, Appli
648	13.0	6	439	22	US-09-823-038A-32	Sequence 32, Appl
649	13.0	6	442	1	PCT-US00-03987-11406	Sequence 11406, A
650	13.0	6	442	1	PCT-US98-06371-1014	Sequence 1014, Ap
651	13.0	6	442	13	US-08-902-615A-280	Sequence 280, App
652	13.0	6	442	16	US-09-270-767-7046	Sequence 7046, Ap
653	13.0	6	442	22	US-09-815-242-11406	Sequence 11406, A
654	13.0	6	444	24	US-10-072-851-11406	Sequence 11406, A
655	13.0	6	444	16	US-09-270-767-33740	Sequence 33740, A
656	13.0	6	444	16	US-09-270-767-48957	Sequence 48957, A
657	13.0	6	446	1	PCT-US99-05606-51	Sequence 51, Appl
658	13.0	6	446	20	US-09-631-275-51	Sequence 51, Appl
659	13.0	6	446	26	US-60-657-474-51	Sequence 51, Appl
660	13.0	6	447	21	US-09-708-427-1527	Sequence 1527, Ap
661	13.0	6	447	26	US-60-208-129-190	Sequence 190, App
662	13.0	6	447	26	US-60-208-973-233	Sequence 233, App
663	13.0	6	448	22	US-09-815-108-6	Sequence 6, Appli
664	13.0	6	449	21	US-09-708-427-2162	Sequence 2162, Ap

us-09-833-017b-2.oligo.rapm

Wed Nov 13 14:19:04 2002

No.					Reference	Sequence
665	6	13.0	450	16	US-09-270-767-46295	Sequence 46295, A
666	6	13.0	451	1	PCT-US01-14827-10545	Sequence 10545, A
667	6	13.0	451	17	US-09-328-352-7659	Sequence 7659, Ap
668	6	13.0	452	15	US-09-126-642-8	Sequence 8, Appli
669	6	13.0	452	18	US-09-463-712C-8	Sequence 8, Appli
670	6	13.0	452	20	US-09-965-415-6	Sequence 6, Appli
671	6	13.0	455	20	US-09-602-874C-54	Sequence 54, Appl
672	6	13.0	455	21	US-09-620-394B-6631	Sequence 6631, Ap
673	6	13.0	455	21	US-09-738-626-4477	Sequence 4477, Ap
674	6	13.0	456	19	US-09-557-262-12	Sequence 12, Appl
675	6	13.0	463	1	PCT-US97-05223-1068	Sequence 1068, Ap
676	6	13.0	463	12	US-08-824-132-1068	Sequence 1068, Ap
677	6	13.0	463	13	US-08-993-002A-5000	Sequence 5000, Ap
678	6	13.0	465	1	PCT-US01-08631-33517	Sequence 33517, A
679	6	13.0	466	1	PCT-US99-05606-64	Sequence 64, Appl
680	6	13.0	466	20	US-09-620-394B-6630	Sequence 6630, Ap
681	6	13.0	466	20	US-09-631-275-64	Sequence 64, Appl
682	6	13.0	466	20	US-09-657-474-64	Sequence 64, Appl
683	6	13.0	467	20	US-09-428-944-2754	Sequence 2754, Ap
684	6	13.0	468	21	US-09-708-427-1526	Sequence 1526, Ap
685	6	13.0	472	22	US-09-815-108-5	Sequence 5, Appli
686	6	13.0	473	21	US-09-708-427-31892	Sequence 31892, A
687	6	13.0	477	1	PCT-US99-05606-4	Sequence 4, Appli
688	6	13.0	477	16	US-09-236-134-4	Sequence 4, Appli
689	6	13.0	477	16	US-09-236-134A-4	Sequence 4, Appli
690	6	13.0	477	16	US-09-236-134B-4	Sequence 4, Appli
691	6	13.0	477	16	US-09-236-134C-4	Sequence 4, Appli
692	6	13.0	477	20	US-09-631-275-4	Sequence 4, Appli
693	6	13.0	477	20	US-09-657-474-4	Sequence 4, Appli
694	6	13.0	477	23	US-09-722-544A-4	Sequence 4, Appli
695	6	13.0	477	23	US-09-971-309-19	Sequence 19, Appl
696	6	13.0	478	26	US-09-614-150-3780	Sequence 3780, Ap
697	6	13.0	478	26	US-60-191-637-3792	Sequence 3792, Ap
698	6	13.0	478	26	US-60-191-681-3048	Sequence 3048, Ap
699	6	13.0	479	19	US-09-583-110-4687	Sequence 4687, Ap
700	6	13.0	480	1	PCT-US00-18198-102	Sequence 102, App
701	6	13.0	480	15	US-09-107-433-3418	Sequence 3418, Ap
702	6	13.0	480	17	US-09-345-464-102	Sequence 102, App
703	6	13.0	480	20	US-09-606-317-102	Sequence 102, App
704	6	13.0	483	16	US-09-252-991A-20139	Sequence 20139, A
705	6	13.0	484	17	US-09-389-956-12	Sequence 12, Appl
706	6	13.0	485	16	US-09-270-767-43385	Sequence 43385, A
707	6	13.0	488	26	US-60-213-178-588	Sequence 588, App
708	6	13.0	489	17	US-09-417-507-29712	Sequence 29712, A
709	6	13.0	489	20	US-09-620-394B-6629	Sequence 6629, Ap
710	6	13.0	491	13	US-08-993-002A-8529	Sequence 8529, Ap
711	6	13.0	493	13	US-09-918-951-5	Sequence 5, Appli
712	6	13.0	494	18	US-09-417-507-42209	Sequence 42209, A
713	6	13.0	494	1	US-09-631-275-133	Sequence 133, App
714	6	13.0	495	1	PCT-US99-05606-2	Sequence 2, Appli
715	6	13.0	495	16	US-09-236-134-2	Sequence 2, Appli
716	6	13.0	495	16	US-09-236-134A-2	Sequence 2, Appli
717	6	13.0	495	16	US-09-236-134B-2	Sequence 2, Appli
718	6	13.0	495	16	US-09-236-134C-2	Sequence 2, Appli
719	6	13.0	495	20	US-09-631-275-2	Sequence 2, Appli
720	6	13.0	495	20	US-09-657-474-2	Sequence 2, Appli
721	6	13.0	495	23	US-09-722-544A-2	Sequence 2, Appli
722	6	13.0	496	16	US-09-201-228A-423	Sequence 423, App
723	6	13.0	496	22	US-09-808-483-14	Sequence 14, Appl
724	6	13.0	499	26	US-60-213-177-791	Sequence 791, App
725	6	13.0	501	26	US-09-614-150-21603	Sequence 21603, A
726	6	13.0	501	26	US-60-167-217-21814	Sequence 21814, A
727	6	13.0	501	26	US-60-173-464-17759	Sequence 17759, A
728	6	13.0	501	26	US-60-191-681-21668	Sequence 21668, A
729	6	13.0	501	26	US-60-191-637-17093	Sequence 17093, A
730	6	13.0	502	15	US-09-134-000-6324	Sequence 6324, Ap
731	6	13.0	504	1	PCT-US00-18198-5	Sequence 5, Appli
732	6	13.0	504	1	PCT-US01-43782-3	Sequence 3, Appli
733	6	13.0	504	1	PCT-US99-08526-2	Sequence 2, Appli
734	6	13.0	504	16	US-09-293-182-2	Sequence 2, Appli
735	6	13.0	504	17	US-09-345-464-5	Sequence 5, Appli
736	6	13.0	504	17	US-09-388-778-2	Sequence 2, Appli
737	6	13.0	504	20	US-09-606-317-5	Sequence 5, Appli
738	6	13.0	504	21	US-09-709-238-119	Sequence 119, App
739	6	13.0	504	21	US-09-758-386-2	Sequence 2, Appli
740	6	13.0	504	21	US-09-796-753-94	Sequence 94, Appl
741	6	13.0	504	21	US-09-796-753-108	Sequence 108, App
742	6	13.0	504	22	US-09-815-108-8	Sequence 8, Appli
743	6	13.0	504	22	US-09-815-108-15	Sequence 15, Appl
744	6	13.0	504	22	US-09-815-108-17	Sequence 17, Appl
745	6	13.0	504	22	US-09-815-108-19	Sequence 19, Appl
746	6	13.0	504	23	US-09-933-480-2	Sequence 2, Appli
747	6	13.0	504	23	US-09-941-992-119	Sequence 119, App
748	6	13.0	504	23	US-09-989-279-119	Sequence 119, App
749	6	13.0	504	23	US-09-989-293A-119	Sequence 119, App
750	6	13.0	504	23	US-09-989-721-119	Sequence 119, App
751	6	13.0	504	23	US-09-989-722-119	Sequence 119, App
752	6	13.0	504	23	US-09-989-723-119	Sequence 119, App
753	6	13.0	504	23	US-09-989-724-119	Sequence 119, App
754	6	13.0	504	23	US-09-989-725-119	Sequence 119, App
755	6	13.0	504	23	US-09-989-726-119	Sequence 119, App
756	6	13.0	504	23	US-09-989-727-119	Sequence 119, App
757	6	13.0	504	23	US-09-989-728-119	Sequence 119, App
758	6	13.0	504	23	US-09-989-729A-119	Sequence 119, App
759	6	13.0	504	23	US-09-989-730-119	Sequence 119, App
760	6	13.0	504	23	US-09-989-731-119	Sequence 119, App
761	6	13.0	504	23	US-09-989-732-119	Sequence 119, App
762	6	13.0	504	23	US-09-989-734-119	Sequence 119, App
763	6	13.0	504	23	US-09-989-735-119	Sequence 119, App
764	6	13.0	504	23	US-09-989-862-119	Sequence 119, App
765	6	13.0	504	23	US-09-990-427-119	Sequence 119, App
766	6	13.0	504	23	US-09-990-436-119	Sequence 119, App
767	6	13.0	504	23	US-09-990-437-119	Sequence 119, App
768	6	13.0	504	23	US-09-990-438-119	Sequence 119, App
769	6	13.0	504	23	US-09-990-439-119	Sequence 119, App
770	6	13.0	504	23	US-09-990-440-119	Sequence 119, App
771	6	13.0	504	23	US-09-990-441-119	Sequence 119, App
772	6	13.0	504	23	US-09-990-442-119	Sequence 119, App
773	6	13.0	504	23	US-09-990-443-119	Sequence 119, App
774	6	13.0	504	23	US-09-990-444-119	Sequence 119, App
775	6	13.0	504	23	US-09-990-456-119	Sequence 119, App
776	6	13.0	504	23	US-09-990-562-119	Sequence 119, App
777	6	13.0	504	23	US-09-990-711-119	Sequence 119, App
778	6	13.0	504	23	US-09-990-726-119	Sequence 119, App
779	6	13.0	504	23	US-09-991-073-119	Sequence 119, App
780	6	13.0	504	23	US-09-991-157-119	Sequence 119, App
781	6	13.0	504	23	US-09-991-163-119	Sequence 119, App
782	6	13.0	504	23	US-09-991-172-119	Sequence 119, App
783	6	13.0	504	23	US-09-991-181-119	Sequence 119, App
784	6	13.0	504	23	US-09-991-854-119	Sequence 119, App
785	6	13.0	504	23	US-09-992-521-119	Sequence 119, App
786	6	13.0	504	23	US-09-992-598-119	Sequence 119, App
787	6	13.0	504	23	US-09-993-469-119	Sequence 119, App
788	6	13.0	504	23	US-09-993-583-119	Sequence 119, App
789	6	13.0	504	23	US-09-993-604-119	Sequence 119, App
790	6	13.0	504	23	US-09-993-667-119	Sequence 119, App
791	6	13.0	504	23	US-09-993-687-119	Sequence 119, App
792	6	13.0	504	23	US-09-993-748-119	Sequence 119, App
793	6	13.0	504	23	US-09-994-054-119	Sequence 119, App
794	6	13.0	504	23	US-09-996-243-119	Sequence 119, App
795	6	13.0	504	23	US-09-997-333-119	Sequence 119, App
796	6	13.0	504	23	US-09-997-349-119	Sequence 119, App
797	6	13.0	504	23	US-09-997-384-119	Sequence 119, App
798	6	13.0	504	23	US-09-997-428-119	Sequence 119, App
799	6	13.0	504	23	US-09-997-440-119	Sequence 119, App
800	6	13.0	504	23	US-09-997-514-119	Sequence 119, App
801	6	13.0	504	23	US-09-997-529-119	Sequence 119, App
802	6	13.0	504	23	US-09-997-542-119	Sequence 119, App
803	6	13.0	504	23	US-09-997-559-119	Sequence 119, App
804	6	13.0	504	23	US-09-997-573-119	Sequence 119, App
805	6	13.0	504	23	US-09-997-585-119	Sequence 119, App
806	6	13.0	504	23	US-09-997-601-119	Sequence 119, App
807	6	13.0	504	23	US-09-997-614-119	Sequence 119, App
808	6	13.0	504	23	US-09-997-628-119	Sequence 119, App
809	6	13.0	504	23	US-09-997-641-119	Sequence 119, App
810	6	13.0	504	23	US-09-997-653-119	Sequence 119, App

Copyright GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2002, 11:02:02 ; Search time 17.8507 Seconds
                                       (without alignments)
                                       247.615 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 46
Sequence: 1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size :   0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   PIR.71:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9	19.6	971	2	D70128	conserved hypothet
2	7	15.2	88	2	G90517	hypothetical prote
3	7	15.2	230	2	S46737	hypothetical prote
4	7	15.2	240	2	F83488	probable ATP-bindi
5	7	15.2	240	2	H95873	probable amino aci
6	7	15.2	240	2	G98155	probable ATP-bindi
7	7	15.2	240	2	AE3132	hypothetical prote
8	7	15.2	253	2	B70483	ABC transporter -
9	7	15.2	253	2	S70367	stem cell factor s
10	7	15.2	287	2	S70366	stem cell factor 1
11	7	15.2	370	2	T48633	hypothetical prote
12	7	15.2	411	1	I55604	platelet glycoprot
13	7	15.2	437	2	AC1112	PTS beta-glucoside
14	7	15.2	437	2	AG1473	PTS beta-glucoside
15	7	15.2	472	2	AD2284	hypothetical prote
16	7	15.2	494	2	B96534	hypothetical prote
17	7	15.2	558	2	A42023	transcription acti
18	7	15.2	565	2	AI0479	probable membrane
19	7	15.2	579	2	A42617	66K rel-related pr
20	7	15.2	591	2	S51303	hypothetical prote
21	7	15.2	618	2	D86364	hypothetical prote
22	7	15.2	788	2	T07667	UDPglucose--glycog
23	7	15.2	808	2	D81924	hypothetical prote
24	7	15.2	1183	2	F90559	conserved hypothet
25	7	15.2	1358	2	T22695	hypothetical prote
26	7	15.2	1804	2	AF2250	serine/threonine k
27	7	15.2	3972	2	S75251	hypothetical prote
28	6	13.0	42	2	A31918	cathepsin D (EC 3.
29	6	13.0	43	2	A97416	hypothetical prote
30	6	13.0	48	2	B82592	hypothetical prote
31	6	13.0	69	2	T18094	hypothetical prote
32	6	13.0	88	2	QXDO	26S proteasome reg
33	6	13.0	97	2	H96743	probable DnaJ prot
34	6	13.0	102	2	G71137	hypothetical prote
35	6	13.0	104	2	B81045	hypothetical prote
36	6	13.0	106	2	AE1842	multidrug exporter
37	6	13.0	106	2	D97574	hypothetical prote
38	6	13.0	106	2	AD2795	conserved hypothet
39	6	13.0	110	2	S73121	cytochrome c553 -
40	6	13.0	114	1	M3HUWE	Ig heavy chain V-I
41	6	13.0	122	1	A1HUTR	Ig heavy chain V-I
42	6	13.0	122	2	B75387	hypothetical prote
43	6	13.0	127	2	A82324	hypothetical prote
44	6	13.0	128	2	AE1330	hypothetical prote
45	6	13.0	133	2	H87681	hypothetical prote
46	6	13.0	134	2	T17655	hypothetical prote
47	6	13.0	141	2	H86609	kinase [imported]
48	6	13.0	152	2	AD1071	conserved hypothet
49	6	13.0	154	2	G86699	hypothetical prote
50	6	13.0	155	2	B70370	hypothetical prote
51	6	13.0	159	2	AI0772	probable exported
52	6	13.0	181	2	S78572	EST3 protein - yea
53	6	13.0	184	2	AD0207	probable exported
54	6	13.0	186	2	A56956	GTP-binding protei
55	6	13.0	197	2	A44363	voltage-gated dihy
56	6	13.0	202	2	H81731	CDP-diacylglycerol
57	6	13.0	202	2	JX0228	multicatalytic end
58	6	13.0	202	2	I49121	multicatalytic end
59	6	13.0	203	2	C95120	hypothetical prote
60	6	13.0	203	2	G97989	hypothetical prote
61	6	13.0	205	2	S17522	multicatalytic end
62	6	13.0	209	2	A55632	proteasome LMP2.s
63	6	13.0	213	2	JE0101	proteasome subunit
64	6	13.0	213	2	E71511	hypothetical prote
65	6	13.0	214	2	G70348	conserved hypothet
66	6	13.0	216	1	XXSACC	chloramphenicol O-
67	6	13.0	219	2	JX0231	proteasome ring12
68	6	13.0	219	2	JC2019	multicatalytic end
69	6	13.0	219	2	S27332	proteasome chain L
70	6	13.0	220	2	B64681	nicotinamide monon
71	6	13.0	220	2	B71832	nicotinamide monon
72	6	13.0	220	2	E98293	hypothetical prote
73	6	13.0	220	2	AD2990	conserved hypothet
74	6	13.0	224	2	T36291	probable glycine-r
75	6	13.0	226	2	T40487	proteasome compone
76	6	13.0	227	2	S66482	transcription regu
77	6	13.0	229	2	S66342	ribonuclease II (E
78	6	13.0	232	2	A99554	hypothetical prote
79	6	13.0	233	2	T04497	multicatalytic end
80	6	13.0	234	2	T03985	multicatalytic end
81	6	13.0	234	2	E84177	hypothetical prote
82	6	13.0	235	2	C82155	probable C-factor
83	6	13.0	236	2	T50209	translin-like prot
84	6	13.0	237	2	A64352	hypothetical prote
85	6	13.0	239	2	B54589	proteasome subunit
86	6	13.0	243	2	D72098	serine esterase, p
87	6	13.0	243	2	G86524	lysophospholipase
88	6	13.0	243	2	C95295	probable ABC trans
89	6	13.0	256	2	F87707	ABC transporter, A
90	6	13.0	257	2	H75199	methylthioadenosin
91	6	13.0	260	2	C71233	probable 5'-methyl
92	6	13.0	262	2	G82901	conserved hypothet
93	6	13.0	265	2	G84743	hypothetical prote
94	6	13.0	270	2	G90404	hemA concentration
95	6	13.0	271	2	G84030	diaminopimelate ep
96	6	13.0	273	2	F64590	diaminopimelate ep
97	6	13.0	273	2	G71923	hypothetical prote
98	6	13.0	274	2	A83981	5'-methylthioadeno
99	6	13.0	274	2	A11729	gp17 (bacteriophag
100	6	13.0	276	2	D41044	octopine/nucleoside
101	6	13.0	277	2	C70351	purine nucleoside
102	6	13.0	278	2	B97069	anaerobic Cobalt c

Left panel

No.	No.	Len	Tm	Accession	Type	Pos	Description
103	176	6	13.0	C23660	2	283	calcium channel pr
104	177	6	13.0	T29047	2	285	hypothetical prote
105	178	6	13.0	E83795	2	285	ABC transporter (A
106	179	6	13.0	JN0637	2	287	stem cell factor p
107	180	6	13.0	D97241	2	287	phosphohydrolase f
108	181	6	13.0	D82578	2	288	conserved hypothet
109	182	6	13.0	T28922	2	289	hypothetical prote
110	183	6	13.0	D90256	2	289	ribose phosphate p
111	184	6	13.0	A23660	2	294	calcium channel pr
112	185	6	13.0	B23660	2	294	calcium channel pr
113	186	6	13.0	S05006	2	295	calcium channel al
114	187	6	13.0	F98205	2	295	probable transcrip
115	188	6	13.0	AC3081	2	295	transcription regu
116	189	6	13.0	A36366	2	296	enhancer-binding p
117	190	6	13.0	A35914	2	297	transcription fact
118	191	6	13.0	E97787	2	297	oxidoreductase Atu
119	192	6	13.0	AF2605	2	297	modulation of CheA
120	193	6	13.0	H83863	2	298	probable transcrip
121	194	6	13.0	E83388	2	300	hypothetical prote
122	195	6	13.0	H98242	2	303	dihydrodipicolinat
123	196	6	13.0	AD3043	2	303	probable oligopept
124	197	6	13.0	C71163	2	304	conserved hypothet
125	198	6	13.0	B82873	2	305	oligopeptide trans
126	199	6	13.0	B75142	2	306	trbG protein - Ent
127	200	6	13.0	T08504	2	306	abc transporter PA
128	201	6	13.0	G75068	2	309	gamma-hemolysin II
129	202	6	13.0	JC4282	2	309	gamma-hemolysin ch
130	203	6	13.0	D90043	2	309	oligopeptide trans
131	204	6	13.0	D64184	2	311	hypothetical prote
132	205	6	13.0	T24625	2	312	probable site-spec
133	206	6	13.0	T45010	2	314	methionyl-tRNA for
134	207	6	13.0	F84987	2	314	ct007 hypothetical
135	208	6	13.0	G72077	2	316	CT007 hypothetical
136	209	6	13.0	G86545	1	316	iron transport pro
137	210	6	13.0	B41671	2	317	hlgA-like protein
138	211	6	13.0	S49271	2	321	hypothetical prote
139	212	6	13.0	H90019	2	323	hypothetical prote
140	213	6	13.0	S76441	2	326	conserved hypothet
141	214	6	13.0	D71316	1	329	hypothetical prote
142	215	6	13.0	I64160	2	329	hypothetical prote
143	216	6	13.0	C75065	2	330	hypothetical prote
144	217	6	13.0	A97556	2	338	pyridine nucleotid
145	218	6	13.0	AC2776	2	338	hypothetical prote
146	219	6	13.0	T42995	2	340	hypothetical prote
147	220	6	13.0	A71957	2	342	flagellar P-ring p
148	221	6	13.0	F64550	2	342	flagellar basal-bo
149	222	6	13.0	KHPGD	1	344	cathepsin D (EC 3.
150	223	6	13.0	D70576	2	344	probable fructose
151	224	6	13.0	S12788	2	345	transcription fact
152	225	6	13.0	S64003	2	349	hypothetical prote
153	226	6	13.0	G64648	1	354	flagellar motor sw
154	227	6	13.0	G71938	2	354	flagellar motor sw
155	228	6	13.0	B86531	2	360	Yop translocation
156	229	6	13.0	C72091	2	360	type III secretion
157	230	6	13.0	F89936	2	366	conserved hypothet
158	231	6	13.0	T22708	2	369	hypothetical prote
159	232	6	13.0	H95173	2	370	hypothetical prote
160	233	6	13.0	A99040	2	370	hypothetical prote
161	234	6	13.0	A46138	2	372	invasion protein i
162	235	6	13.0	AE0852	2	372	cell invasion prot
163	236	6	13.0	T09374	2	375	hypothetical prote
164	237	6	13.0	S73530	2	375	MG438 homolog E30_
165	238	6	13.0	JC2142	1	379	alkaline proteinas
166	239	6	13.0	G96804	2	381	hypothetical prote
167	240	6	13.0	AF2427	2	385	aspartic proteinas
168	241	6	13.0	A45117	2	387	hypothetical prote
169	242	6	13.0	T25590	2	393	aspartic proteinas
170	243	6	13.0	AB0073	2	395	probable methyltra
171	244	6	13.0	T47207	2	396	aspartic proteinas
172	245	6	13.0	AE3337	2	401	oligopeptide trans
173	246	6	13.0	A85832	2	405	probable glycosyl
174	247	6	13.0	H64971	2	405	probable colanic a
175	248	6	13.0	F90986	2	405	probable glycosyl

Right panel

Index	Len	Tm	Accession	Type	Pos	Description
407	6	13.0	KHRTD	1	407	cathepsin D (EC 3.
410	6	13.0	KHMSD	1	410	cathepsin D (EC 3.
412	6	13.0	KHHUD	2	412	cathepsin D (EC 3.
416	6	13.0	AC1143	2	416	probable NAD(P)-de
416	6	13.0	AF1501	2	416	probable NAD(P)-de
418	6	13.0	B81359	2	418	UDP-N-acetylglucos
418	6	13.0	G69302	2	418	sulfite reductase
421	6	13.0	C95880	2	421	probable protein [
422	6	13.0	T33374	2	422	hypothetical prote
427	6	13.0	F75406	2	427	histidyl-tRNA synt
428	6	13.0	T31746	2	428	hypothetical prote
429	6	13.0	S59773	2	429	26S proteasome reg
433	6	13.0	H97199	2	433	htrA-like serine p
434	6	13.0	E97026	2	434	aspartate oxidase
442	6	13.0	F64668	1	442	histidine--tRNA li
442	6	13.0	D71847	2	442	histidine--tRNA li
447	6	13.0	G84687	2	447	probable disease r
457	6	13.0	H85013	2	457	hypothetical prote
460	6	13.0	AG0965	2	460	probable glycosyl
468	6	13.0	T52269	2	468	1,2-diacylglycerol
472	6	13.0	T41684	2	472	probable sterol o-
476	6	13.0	A41463	2	476	alkaline metallopr
478	6	13.0	H71224	2	478	hypothetical prote
479	6	13.0	S26699	2	479	alkaline metallopr
479	6	13.0	E95055	2	479	potassium uptake p
481	6	13.0	G75192	2	481	probable ss-DNA-sp
484	6	13.0	T50909	2	484	hypothetical prote
484	6	13.0	S61006	2	484	RIO1 protein - yea
493	6	13.0	H70573	2	493	hypothetical prote
494	6	13.0	S48769	2	494	hypothetical prote
497	6	13.0	D69853	2	497	altronate hydrolas
499	6	13.0	DELVN4	1	499	NADH dehydrogenase
499	6	13.0	S39113	2	499	alpha-L-arabinofur
506	6	13.0	C64377	2	506	hypothetical prote
508	6	13.0	T22440	2	508	hypothetical prote
518	6	13.0	H69649	2	518	2-isopropylmalate
519	6	13.0	T24243	2	519	hypothetical prote
521	6	13.0	S08435	2	521	gag polyprotein -
530	6	13.0	T43089	2	530	transfer complex p
538	6	13.0	I51560	2	538	ribonucleoprotein
550	6	13.0	A42679	1	550	protein-tyrosine-p
560	6	13.0	D71913	2	560	lytic murein trans
560	6	13.0	E64600	2	560	soluble lytic mure
561	6	13.0	E64980	1	561	hypothetical 62.1
561	6	13.0	A96996	2	561	probable 2-compone
561	6	13.0	F85841	2	561	probable 2-compone
561	6	13.0	AE0777	2	561	probable two-compo
567	6	13.0	E84658	2	567	hypothetical prote
569	6	13.0	G81400	2	569	proline--tRNA liga
569	6	13.0	S11035	2	569	chaperonin hsp60,
571	6	13.0	C83527	2	571	prolyl-tRNA synthe
572	6	13.0	C42679	2	572	protein-tyrosine-p
573	6	13.0	A32800	2	573	chaperonin GroEL p
573	6	13.0	HHRT60	2	573	chaperonin groEL p
573	6	13.0	HHMS60	1	573	chaperonin groEL p
575	6	13.0	A34173	2	575	mitochondrial prot
575	6	13.0	E88651	2	575	protein B0212.2 [1
579	6	13.0	T47705	2	579	hypothetical prote
583	6	13.0	T15644	2	583	protein Rl9.7 [im
593	6	13.0	G87719	2	593	penicillin-binding
594	6	13.0	D97806	2	594	penicillin-binding
599	6	13.0	C71661	2	599	protein-tyrosine-p
604	6	13.0	I51405	2	604	prolyl-tRNA synthe
612	6	13.0	AE2437	2	612	hypothetical prote
612	6	13.0	T50226	2	612	hypothetical prote
617	6	13.0	T29761	2	617	proline-tRNA ligas
619	6	13.0	S75447	2	619	taurine transport
620	6	13.0	S29839	2	620	taurine transporte
620	6	13.0	S46487	2	620	taurine transporte
634	6	13.0	G01426	2	634	probable permease
638	6	13.0	E96927	2	638	methylmalonyl-CoA
638	6	13.0	S04640	2	638	probable permease
639	6	13.0	E72336	2	639	maltodextrin glyco

#					ID	Description
249	6	13.0	639	2	S20887	actA protein precu
250	6	13.0	639	2	AE1100	actin-assembly ind
251	6	13.0	649	2	H90163	DNA-directed RNA p
252	6	13.0	662	2	S62707	NADH dehydrogenase
253	6	13.0	664	2	A72215	excinuclease ABC c
254	6	13.0	671	2	T33491	hypothetical prote
255	6	13.0	672	2	AI1183	transcription regu
256	6	13.0	688	2	AF1619	glycyl-tRNA synthe
257	6	13.0	699	2	T15920	hypothetical prote
258	6	13.0	702	2	E69498	hypothetical prote
259	6	13.0	715	2	T25531	hypothetical prote
260	6	13.0	717	1	S28784	alpha-amylase (EC
261	6	13.0	721	2	F70649	probable Acyl-CoA
262	6	13.0	726	2	A87334	ABC transporter, H
263	6	13.0	735	2	I50630	alpha subunit of c
264	6	13.0	737	2	S72442	actin-fragmin kina
265	6	13.0	758	2	AE2022	hypothetical prote
266	6	13.0	759	2	S53922	PMT6 protein - yea
267	6	13.0	804	2	T37821	probable dna repai
268	6	13.0	805	2	T21592	hypothetical prote
269	6	13.0	817	2	T22442	hypothetical prote
270	6	13.0	825	2	T27852	hypothetical prote
271	6	13.0	829	2	S12858	virA protein - Agr
272	6	13.0	830	2	A84951	DNA topoisomerase
273	6	13.0	831	2	D96639	protein TIF9.12 [i
274	6	13.0	833	2	S06972	virA protein - Agr
275	6	13.0	833	2	AD3248	two component sens
276	6	13.0	841	2	A90669	probable enzyme (i
277	6	13.0	841	2	C64755	yagX protein - Esc
278	6	13.0	841	2	D85519	probable enzyme ya
279	6	13.0	842	2	T23715	hypothetical prote
280	6	13.0	857	2	AH2768	protein-export mem
281	6	13.0	874	2	G84644	probable ATP-depen
282	6	13.0	875	2	H90371	proteinase [import
283	6	13.0	876	2	B97549	probable protein-e
284	6	13.0	879	2	S73757	hypothetical prote
285	6	13.0	887	1	RDHYE	hydroxymethylgluta
286	6	13.0	887	2	A23586	hydroxymethylgluta
287	6	13.0	888	1	RDHUE	hydroxymethylgluta
288	6	13.0	890	2	S44150	coat protein - str
289	6	13.0	893	2	E72232	DNA-directed DNA p
290	6	13.0	968	2	T01733	hypothetical prote
291	6	13.0	970	2	G84939	hypothetical prote
292	6	13.0	971	2	T33907	hypothetical prote
293	6	13.0	977	2	B64593	hypothetical prote
294	6	13.0	980	2	AD1922	methyl-accepting c
295	6	13.0	1012	2	T43054	retinoblastoma-rel
296	6	13.0	1032	2	T23164	hypothetical prote
297	6	13.0	1035	2	T23165	hypothetical prote
298	6	13.0	1049	2	T37953	hypothetical prote
299	6	13.0	1052	2	T51904	hypothetical prote
300	6	13.0	1052	2	AF2959	conserved hypothet
301	6	13.0	1054	2	D96519	mysoin-like protei
302	6	13.0	1072	2	G95851	probable hemolysin
303	6	13.0	1075	2	T38253	hypothetical prote
304	6	13.0	1110	2	I51116	NF-180 - sea lampr
305	6	13.0	1126	2	S04716	DNA-directed RNA p
306	6	13.0	1160	2	T23713	hypothetical prote
307	6	13.0	1164	2	G71827	hypothetical prote
308	6	13.0	1184	2	A96638	dna-directed RNA p
309	6	13.0	1286	2	T23714	hypothetical prote
310	6	13.0	1291	2	T17242	hypothetical prote
311	6	13.0	1331	2	S05011	calcium channel al
312	6	13.0	1341	2	H98323	hypothetical prote
313	6	13.0	1355	2	A47026	alkaline phosphata
314	6	13.0	1373	2	E97722	hypothetical prote
315	6	13.0	1374	2	A71724	dna-directed RNA p
316	6	13.0	1529	2	T02730	RNA-directed DNA p
317	6	13.0	1559	2	T30535	calcium channel al
318	6	13.0	1610	2	A46227	voltage-dependent
319	6	13.0	1642	2	T08880	NMDA receptor-bind
320	6	13.0	1646	2	JH0422	voltage-dependent
321	6	13.0	1748	1	JN0786	integrin beta-4 ch
322	6	13.0	1787	2	G81684	excinuclease ABC c
323	6	13.0	1807	2	JC6319	integrin beta-4 ch
324	6	13.0	1852	2	A37860	calcium channel pr
325	6	13.0	1855	2	S41649	DNA polymerase - m
326	6	13.0	1864	2	F86378	protein F2109.12 [
327	6	13.0	1946	2	AC2141	serine/threonine k
328	6	13.0	2022	2	T48818	glucan 1,4-alpha-g
329	6	13.0	2100	2	T38128	t7123.15 protein -
330	6	13.0	2139	2	A44467	voltage-dependent
331	6	13.0	2143	2	JH0427	voltage-dependent
332	6	13.0	2161	2	JH0564	calcium channel al
333	6	13.0	2166	2	S11339	calcium channel pr
334	6	13.0	2171	2	S05054	calcium channel al
335	6	13.0	2181	2	A38198	calcium channel al
336	6	13.0	2203	2	T42742	voltage-dependent
337	6	13.0	2220	2	A45290	calcium channel pr
338	6	13.0	2271	2	F90073	hypothetical prote
339	6	13.0	2425	2	D69426	surface layer prot
340	6	13.0	2693	2	A40743	IP3 receptor, XIP3
341	6	13.0	2747	2	B49132	fat facets (faf) s
342	6	13.0	2946	2	T15840	hypothetical prote
343	6	13.0	3194	2	D71917	toxin-like outer m
344	6	13.0	3512	2	T17121	CPY protein - midg
345	6	13.0	4910	2	S64942	probable membrane
346	6	13.0	8243	2	T31307	type I fatty acid
347	6	13.0	13288	2	T03099	mucin, submaxillar
348	1	13.0	26926	1	I38344	titin, cardiac mus
349	6	10.9	24	2	PN0653	alkaline trypsin-1
350	5	10.9	26	2	S04376	hypothetical prote
351	5	10.9	36	2	B44400	myosin heavy chain
352	5	10.9	36	2	H70243	hypothetical prote
353	5	10.9	42	2	JE0005	hypothetical prote
354	5	10.9	42	2	A99578	actin alpha, smoot
355	5	10.9	44	2	C47193	lipoprotein vsaH [
356	5	10.9	45	1	HPBO	1-caldesmon I (HeL
357	5	10.9	46	2	S42240	haptoglobin precur
358	5	10.9	46	2	I38225	hypothetical prote
359	5	10.9	46	2	A45758	protein-serine/thr
360	5	10.9	47	1	D81117	hypothetical gluco
361	5	10.9	48	2	B64647	hypothetical prote
362	5	10.9	50	2	F70227	hypothetical prote
363	5	10.9	51	2	A70229	hypothetical prote
364	5	10.9	51	2	G82455	hypothetical prote
365	5	10.9	52	2	D82682	hypothetical prote
366	5	10.9	53	2	S12520	core protein A1 -
367	5	10.9	53	2	F82662	hypothetical prote
368	5	10.9	53	2	AB1896	hypothetical prote
369	5	10.9	54	2	I46468	actin - sheep (fra
370	5	10.9	58	2	E97047	hypothetical prote
371	5	10.9	60	1	LPSE5	lantibiotic Pep5 p
372	5	10.9	62	2	A65045	hypothetical prote
373	5	10.9	62	2	E69346	hypothetical prote
374	5	10.9	63	2	T12825	conserved hypothet
375	5	10.9	64	2	A86333	hypothetical prote
376	5	10.9	65	2	S73176	ribosomal protein
377	5	10.9	66	2	T29601	hypothetical prote
378	5	10.9	67	2	I46473	alpha-actin - rabb
379	5	10.9	67	2	C86830	hypothetical prote
380	5	10.9	67	2	E81135	hypothetical prote
381	5	10.9	67	2	E81042	hypothetical prote
382	5	10.9	67	2	B69447	hypothetical prote
383	5	10.9	67	2	T07256	hypothetical prote
384	5	10.9	69	1	MIEC77	microcin B17 precu
385	5	10.9	72	2	A35511	homeotic protein H
386	5	10.9	72	2	S75519	alpha-actin - rabb
387	5	10.9	73	1	HPRB	haptoglobin precur
388	5	10.9	75	2	H81320	small hydrophobic
389	5	10.9	76	2	B75271	hypothetical prote
390	5	10.9	77	2	T14995	hypothetical prote
391	5	10.9	78	2	T31106	probable glycerol-
392	5	10.9	78	2	C71978	hypothetical prote
393	5	10.9	78	2	S17872	translation initia
394	5	10.9	79	2	H82705	hypothetical prote

#			Seq		Accession	Description
395	10.9	5	79	2	AH3369	hypothetical prote
396	10.9	5	80	2	E71430	hypothetical prote
397	10.9	5	81	2	C97430	hypothetical prote
398	10.9	5	82	2	T18102	hypothetical prote
399	10.9	5	82	2	AC2753	conserved hypothet
400	10.9	5	83	2	A46348	hypothetical prote
401	10.9	5	83	2	A96949	hypothetical prote
402	10.9	5	83	2	F98226	hypothetical prote
403	10.9	5	83	2	AB3060	conserved hypothet
404	10.9	5	83	2	AG3378	hypothetical membr
405	10.9	5	83	4	S53038	hypothetical prote
406	10.9	5	84	2	E90475	hypothetical prote
407	10.9	5	84	2	F84014	hypothetical prote
408	10.9	5	84	2	AF2777	conserved hypothet
409	10.9	5	85	2	T26333	hypothetical prote
410	10.9	5	85	1	A36437	flavocytochrome c,
411	10.9	5	87	2	D71286	probable ribosomal
412	10.9	5	87	2	B97266	hypothetical prote
413	10.9	5	88	2	AC0730	hypothetical prote
414	10.9	5	89	2	XVBBBD	Shigella toxin cha
415	10.9	5	89	2	JN0726	Shiga-like toxin I
416	10.9	5	89	2	XVBPH9	Shiga-like toxin c
417	10.9	5	89	2	I56163	calgranulin A - mo
418	10.9	5	89	2	B53887	Shiga-like toxin I
419	10.9	5	89	2	E91000	Shiga toxin I subu
420	10.9	5	89	2	E83141	hypothetical prote
421	10.9	5	89	2	G85845	Shiga toxin I subu
422	10.9	5	90	2	A54937	brain type III sod
423	10.9	5	90	2	F90555	hypothetical prote
424	10.9	5	91	2	I54248	apolipoprotein A-I
425	10.9	5	91	2	G64028	hypothetical prote
426	10.9	5	92	2	B45837	Ig gamma-1 chain C
427	10.9	5	92	2	I55688	seminal vesicle pr
428	10.9	5	92	2	S57260	nitrite reductase
429	10.9	5	92	2	A81105	probable periplasm
430	10.9	5	92	2	D82451	conserved hypothet
431	10.9	5	92	2	AE2080	hypothetical prote
432	10.9	5	92	2	AE3173	transcription regu
433	10.9	5	93	2	C64684	ribosomal protein
434	10.9	5	93	2	C71835	ribosomal protein
435	10.9	5	93	2	AI0757	probable cobalt tr
436	10.9	5	93	2	AG1359	holin [Bacteriopha
437	10.9	5	93	2	AH1448	holin (Bacteriopha
438	10.9	5	94	2	H42518	A15L 11K protein -
439	10.9	5	94	2	G36849	A15L protein - var
440	10.9	5	94	2	T28557	hypothetical prote
441	10.9	5	94	2	E72165	Al6L protein - var
442	10.9	5	95	2	S70353	KIN28 protein homo
443	10.9	5	95	2	A70803	early secretory an
444	10.9	5	95	2	C69449	conserved hypothet
445	10.9	5	96	2	S69800	holin protein 10.3
446	10.9	5	96	2	S69798	holin protein, 10.
447	10.9	5	97	2	S26652	Ig gamma-1 chain C
448	10.9	5	97	2	E95064	conserved hypothet
449	10.9	5	97	2	C90661	hypothetical prote
450	10.9	5	97	2	A64748	yafN protein - Esc
451	10.9	5	97	2	G64454	hypothetical prote
452	10.9	5	97	2	F84155	bltD operon transc
453	10.9	5	97	2	A85512	hypothetical prote
454	10.9	5	97	2	G97931	conserved hypothet
455	10.9	5	97	2	AH0965	probable secreted
456	10.9	5	98	2	T42307	integral membrane
457	10.9	5	98	2	A87166	hypothetical prote
458	10.9	5	99	2	AC0152	urease operon 23K
459	10.9	5	99	2	AB3452	actin - fruit fly
460	10.9	5	100	2	S14851	conserved hypothet
461	10.9	5	100	2	A87463	hypothetical prote
462	10.9	5	100	2	T49621	H+-transporting AT
463	10.9	5	101	2	T45106	calcium channel, v
464	10.9	5	101	2	I38611	small hydrophobic
465	10.9	5	101	2	H81352	probable membrane
466	10.9	5	102	2	S51974	hypothetical prote
467	10.9	5	102	2	A97534	hypothetical prote
468	10.9	5	102	2	G83721	hypothetical prote
469	10.9	5	102	2	AD1748	B. subtilis YfhL p
470	10.9	5	103	2	A46011	cysteine proteinas
471	10.9	5	103	2	A86584	CT631 hypothetical
472	10.9	5	104	2	S53942	probable membrane
473	10.9	5	104	2	B71450	hypothetical prote
474	10.9	5	104	2	S51479	drought-induced pr
475	10.9	5	104	2	F86867	prophage ps3 prote
476	10.9	5	104	2	AG3518	hypothetical prote
477	10.9	5	105	2	T17976	hypothetical prote
478	10.9	5	105	2	A82914	hypothetical prote
479	10.9	5	105	2	E81963	hypothetical IS110
480	10.9	5	105	2	H32537	T-cell receptor be
481	10.9	5	106	2	A03859	hypothetical prote
482	10.9	5	106	2	H69310	hypothetical prote
483	10.9	5	106	2	D70212	hypothetical prote
484	10.9	5	106	2	G97841	acetyl-CoA acetylt
485	10.9	5	107	2	D71114	hypothetical prote
486	10.9	5	107	2	JC1127	major allergen cha
487	10.9	5	107	2	B85356	glycine-rich prote
488	10.9	5	107	2	E69318	hypothetical prote
489	10.9	5	108	2	T45320	hypothetical prote
490	10.9	5	108	2	D70711	hypothetical prote
491	10.9	5	108	2	F64010	hypothetical prote
492	10.9	5	108	2	AB0218	probable acyl carr
493	10.9	5	109	2	S31978	Ig kappa chain - h
494	10.9	5	109	2	C56413	major allergen Fel
495	10.9	5	109	2	T35249	hypothetical prote
496	10.9	5	110	2	AG3382	quaternary ammoniu
497	10.9	5	110	2	S70789	csgC protein precu
498	10.9	5	110	2	E90806	probable curli pro
499	10.9	5	110	2	A85666	probable curli pro
500	10.9	5	110	2	B64889	ydbL protein precu
501	10.9	5	111	2	S56499	hypothetical 13.1K
502	10.9	5	111	2	T31580	hypothetical prote
503	10.9	5	112	2	T26105	hypothetical prote
504	10.9	5	113	2	JQ1128	gas-vesicle operon
505	10.9	5	113	2	T08234	gas-vesicle operon
506	10.9	5	113	2	A83697	hypothetical prote
507	10.9	5	113	2	I51025	transcription fact
508	10.9	5	113	2	F81370	hypothetical prote
509	10.9	5	113	2	JN0784	acid phosphatase,
510	10.9	5	114	2	F71141	hypothetical prote
511	10.9	5	115	2	T30042	hypothetical prote
512	10.9	5	115	2	T12916	hypothetical prote
513	10.9	5	115	2	E86809	hypothetical prote
514	10.9	5	116	2	H69023	conserved hypothet
515	10.9	5	116	2	A12446	hypothetical prote
516	10.9	5	116	2	C72232	hypothetical prote
517	10.9	5	117	1	A59316	ghrelin precursor
518	10.9	5	117	2	F72620	hypothetical prote
519	10.9	5	117	2	S28124	gas-vesicle operon
520	10.9	5	117	2	JX0053	lysozyme (EC 3.2.1
521	10.9	5	117	2	T49511	hypothetical prote
522	10.9	5	118	2	JC2568	mraY protein - Rhi
523	10.9	5	118	2	T06659	histone H2A.T6G15.
524	10.9	5	118	2	T28246	hypothetical prote
525	10.9	5	118	2	T30509	hypothetical prote
526	10.9	5	119	2	S28133	gas vesicle protei
527	10.9	5	119	2	B90168	conserved hypothet
528	10.9	5	119	2	H70437	hypothetical prote
529	10.9	5	120	2	H84973	ribosome-binding f
530	10.9	5	120	2	F71104	hypothetical prote
531	10.9	5	120	2	B71020	hypothetical prote
532	10.9	5	120	2	G71075	hypothetical prote
533	10.9	5	122	1	A42628	calgranulin B - bo
534	10.9	5	122	2	S69909	Ig V-D-J region (M
535	10.9	5	122	2	T41471	probable ribosomal
536	10.9	5	122	2	S29540	seminal vesicle se
537	10.9	5	122	2	G81855	insertion element
538	10.9	5	122	2	AB1199	hypothetical prote
539	10.9	5	122	2	AC1848	hypothetical prote
540	10.9	5	123	2	D72519	hypothetical prote

Left block (seq 541–613):

Seq	Accession	n	Pos	Tm	Len	Description
541	G72581	2	123	10.9	5	hypothetical prote
542	F83507	2	123	10.9	5	hypothetical prote
543	S03521	2	124	10.9	5	Ig kappa chain pre
544	A26337	2	125	10.9	5	actin, skeletal mu
545	C83138	2	125	10.9	5	hypothetical prote
546	T38515	2	125	10.9	5	hypothetical prote
547	AG1653	2	126	10.9	5	Antigen C homolog
548	S50574	2	128	10.9	5	hypothetical prote
549	AB487	2	128	10.9	5	hypothetical prote
550	T32789	2	128	10.9	5	hypothetical prote
551	AEI701	2	129	10.9	5	hypothetical prote
552	F72488	2	129	10.9	5	hypothetical prote
553	E90535	2	129	10.9	5	hypothetical prote
554	B90175	2	130	10.9	5	NADH dehydrogenase
555	PL0098	2	131	10.9	5	Ig heavy chain pre
556	G71212	2	131	10.9	5	hypothetical prote
557	S67869	2	132	10.9	5	probable membrane
558	S63305	2	133	10.9	5	probable membrane
559	B69489	2	133	10.9	5	LSU ribosomal prot
560	AC3143	2	133	10.9	5	hypothetical prote
561	D69836	2	134	10.9	5	hypothetical prote
562	F71189	2	134	10.9	5	hypothetical prote
563	C97557	2	134	10.9	5	hypothetical prote
564	F75170	2	135	10.9	5	hypothetical prote
565	T43561	2	135	10.9	5	probable transposa
566	S72727	2	135	10.9	5	dedA protein - Myc
567	S38386	2	135	10.9	5	T-cell receptor be
568	T49327	2	136	10.9	5	cofilin related pr
569	G64600	2	136	10.9	5	hypothetical prote
570	G86695	2	137	10.9	5	orf10 protein - Au
571	T31350	2	137	10.9	5	hypothetical prote
572	T02870	2	137	10.9	5	hypothetical prote
573	AG2009	2	138	10.9	5	globulin 2 precurs
574	T33767	2	138	10.9	5	transcription regu
575	A28258	2	138	10.9	5	probable transcrip
576	T04930	2	139	10.9	5	actin 5C - fruit f
577	AB2544	2	140	10.9	5	glycine-rich cell
578	G72666	2	140	10.9	5	transcription regu
579	A69219	1	140	10.9	5	conserved hypothet
580	S67968	2	141	10.9	5	transition protein
581	H72452	2	141	10.9	5	hypothetical prote
582	AEI324	2	141	10.9	5	hypothetical prote
583	A71190	2	142	10.9	5	PTS mannose-specif
584	S73680	2	142	10.9	5	hypothetical prote
585	H69405	2	142	10.9	5	MG337 homolog P02_
586	JQ2293	4	143	10.9	5	hypothetical prote
587	A84457	2	143	10.9	5	transcription enha
588	T48476	2	143	10.9	5	40S ribosomal prot
589	S30832	2	143	10.9	5	40S ribosomal prot
590	F71176	2	143	10.9	5	hypothetical prote
591	B64399	2	143	10.9	5	hypothetical prote
592	T08343	2	143	10.9	5	arsC protein - Hal
593	T46832	2	144	10.9	5	heat shock protein
594	F97079	2	144	10.9	5	PTS system, fructo
595	S17939	2	144	10.9	5	xpsG protein - Xan
596	D72785	2	144	10.9	5	yhcV homolog APE02
597	T88838	2	144	10.9	5	hypothetical prote
598	G72000	2	145	10.9	5	Hth transcription
599	B86624	2	145	10.9	5	HTH transcription
600	T49689	2	145	10.9	5	hypothetical prote
601	JC7309	2	145	10.9	5	agglutinin I - Jer
602	T33494	2	145	10.9	5	hypothetical prote
603	H71853	2	145	10.9	5	hypothetical prote
604	H64659	2	145	10.9	5	hypothetical prote
605	G81910	2	145	10.9	5	probable periplasm
606	AFI554	2	145	10.9	5	B. subtilis YjcF p
607	A47688	2	145	10.9	5	hemagglutinating s
608	T15089	2	145	10.9	5	hypothetical prote
609	A87526	2	145	10.9	5	hypothetical prote
610	I39564	2	145	10.9	5	conserved hypothet
611	C82513	2	145	10.9	5	probable maturase-

Right block (seq 614–686):

Seq	Accession	n	Pos	Tm	Len	Description
614	T06471	2	146	10.9	5	core protein - gar
615	C89989	2	146	10.9	5	conserved hypothet
616	A69315	1	147	10.9	5	proteinase synthas
617	D90126	2	147	10.9	5	ubiquitin-conjugat
618	AH1196	2	147	10.9	5	B. subtilis YjcF p
619	A81454	2	147	10.9	5	probable integral
620	A69554	2	147	10.9	5	conserved hypothet
621	B69040	2	147	10.9	5	hypothetical prote
622	C70312	2	147	10.9	5	hypothetical prote
623	F83801	2	147	10.9	5	chorismate mutase
624	A99145	2	147	10.9	5	hypothetical prote
625	E64370	2	147	10.9	5	hypothetical prote
626	A39091	2	147	10.9	5	phospholipase A2 i
627	B39091	2	147	10.9	5	phospholipase A2 i
628	A75093	2	147	10.9	5	hypothetical prote
629	F71050	2	147	10.9	5	hypothetical prote
630	E64939	2	148	10.9	5	hypothetical prote
631	B90941	2	148	10.9	5	hypothetical prote
632	F85789	2	148	10.9	5	conserved hypothet
633	D87687	2	148	10.9	5	hypothetical prote
634	S10655	2	148	10.9	5	hypothetical prote
635	A96641	2	148	10.9	5	hypothetical prote
636	T52250	2	148	10.9	5	probable transcrip
637	AC1122	2	148	10.9	5	hypothetical prote
638	AI2051	2	148	10.9	5	two-component resp
639	C95188	2	149	10.9	5	PTS system, IIA co
640	H72344	2	149	10.9	5	conserved hypothet
641	D64693	2	149	10.9	5	hypothetical prote
642	C90824	2	149	10.9	5	hypothetical membr
643	MOCHA2	2	150	10.9	5	myosin alkali ligh
644	G85939	1	150	10.9	5	type III secretion
645	MCUR2C	2	151	10.9	5	Spec2c protein - s
646	T42298	1	151	10.9	5	hypothetical prote
647	AD1569	2	151	10.9	5	hypothetical prote
648	AG1215	2	151	10.9	5	hypothetical prote
649	D87548	2	151	10.9	5	hypothetical prote
650	B86474	2	151	10.9	5	hypothetical prote
651	B89851	2	152	10.9	5	hypothetical prote
652	S72685	2	152	10.9	5	hypothetical prote
653	AE2431	2	153	10.9	5	hypothetical prote
654	F90704	2	153	10.9	5	hypothetical prote
655	H72777	2	153	10.9	5	hypothetical prote
656	H83895	2	153	10.9	5	hypothetical prote
657	A85555	2	153	10.9	5	transcription regu
658	JS0710	2	154	10.9	5	hypothetical prote
659	S24396	2	154	10.9	5	heat shock protein
660	T44801	2	154	10.9	5	heat shock protein
661	H82810	2	154	10.9	5	heat shock protein
662	H75031	2	154	10.9	5	bacterioferritin X
663	D83516	2	154	10.9	5	hypothetical prote
664	G89932	2	154	10.9	5	hypothetical prote
665	A60945	2	155	10.9	5	actin homolog FAT,
666	F71004	2	155	10.9	5	hypothetical prote
667	A71693	2	155	10.9	5	hypothetical prote
668	C91094	2	155	10.9	5	type III secretion
669	A49342	2	156	10.9	5	type III secretion
670	D69019	1	156	10.9	5	acetyl-CoA carboxy
671	F86661	2	156	10.9	5	acetyltransferase
672	T16531	2	156	10.9	5	hypothetical prote
673	A71137	2	157	10.9	5	hypothetical prote
674	S75191	2	157	10.9	5	hypothetical prote
675	G97760	2	157	10.9	5	hypothetical prote
676	AG0537	2	157	10.9	5	probable secreted
677	T48178	2	157	10.9	5	hypothetical prote
678	T46440	2	157	10.9	5	hypothetical prote
679	D82311	2	158	10.9	5	conserved hypothet
680	B82191	2	158	10.9	5	hypothetical prote
681	F83475	2	158	10.9	5	conserved hypothet
682	T25554	2	158	10.9	5	hypothetical prote
683	A43599	2	159	10.9	5	cytolysin IIC - Ac
684	S15640	2	159	10.9	5	pilin (variant NGF
685	G82924	2	159	10.9	5	hypothetical UU180
686	D95261	2	159	10.9	5	conserved hypothet

No.		Code				Description
687	2	A98127	159	10.9	5	conserved hypothet
688	2	I55083	159	10.9	5	2c-methyl-D-erythr
689	2	H91078	159	10.9	5	2c-methyl-D-erythr
690	2	A85924	159	10.9	5	2c-methyl-D-erythr
691	2	AD0856	159	10.9	5	2c-methyl-D-erythr
692	2	S15964	159	10.9	5	hypothetical prote
693	2	E95328	159	10.9	5	hypothetical prote
694	2	G88082	159	10.9	5	protein T05A8.1 [i
695	2	G84125	159	10.9	5	hypothetical prote
696	2	B43834	160	10.9	5	RTX hemolysin acti
697	2	H82207	160	10.9	5	hypothetical prote
698	2	T32026	160	10.9	5	actin - shore pine
699	2	S26039	161	10.9	5	hypothetical prote
700	2	B71687	161	10.9	5	hypothetical prote
701	2	G64355	161	10.9	5	cysteine dioxygena
702	2	F70006	161	10.9	5	conserved hypothet
703	2	B90417	161	10.9	5	probable bacteriop
704	2	AD0836	161	10.9	5	probable phage tai
705	2	AD0930	161	10.9	5	F1N18.18 protein -
706	2	B86421	161	10.9	5	hosphoribosylamino
707	2	AE1668	162	10.9	5	hosphoribosylamino
708	2	AG1296	162	10.9	5	glycine-rich prote
709	2	C85356	162	10.9	5	hypothetical prote
710	2	T35903	162	10.9	5	hypothetical prote
711	2	F86637	162	10.9	5	hypothetical prote
712	2	S66764	162	10.9	5	cytochrome c550 [i
713	2	AC1839	162	10.9	5	3-isopropylmalate
714	2	A75160	163	10.9	5	hypothetical prote
715	2	A70369	163	10.9	5	protein disulfide
716	2	G81183	164	10.9	5	hypothetical prote
717	2	I40526	164	10.9	5	nimB protein - Bac
718	2	I40183	164	10.9	5	hypothetical prote
719	2	AC0537	164	10.9	5	hypothetical prote
720	2	T01777	164	10.9	5	calcineurin regula
721	2	PS0261	165	10.9	5	general secretory
722	2	B82670	165	10.9	5	N-acetyltransferas
723	2	AH1249	165	10.9	5	keratin, 68K type
724	1	KRBO2B	166	10.9	5	hypothetical prote
725	2	AE2392	166	10.9	5	hypothetical prote
726	2	AG3566	166	10.9	5	hypothetical prote
727	2	D72394	166	10.9	5	probable membrane
728	2	F97128	166	10.9	5	hypothetical prote
729	2	AI1444	166	10.9	5	probable signal pe
730	2	B81676	167	10.9	5	hypothetical prote
731	2	B71553	167	10.9	5	histone H1 [import
732	2	F84559	167	10.9	5	hypothetical prote
733	2	E64367	168	10.9	5	hypothetical prote
734	2	T25892	168	10.9	5	conserved hypothet
735	2	G82898	168	10.9	5	cytochrome-c biosy
736	2	S54745	169	10.9	5	hypothetical prote
737	2	F82403	169	10.9	5	hypothetical prote
738	2	C84923	169	10.9	5	hypothetical prote
739	2	B87602	169	10.9	5	prophage pi3 prote
740	2	D86801	169	10.9	5	calcineurin regula
741	2	A33391	170	10.9	5	calcineurin regula
742	1	S34127	170	10.9	5	calcineurin regula
743	1	JC1220	170	10.9	5	Ig gamma-2a chain
744	2	A35944	170	10.9	5	calcineurin regula
745	2	JC7242	170	10.9	5	hypothetical prote
746	2	JC2213	170	10.9	5	hypothetical 14.7K
747	2	T27263	170	10.9	5	hypothetical prote
748	2	F82908	170	10.9	5	conserved hypothet
749	2	C90079	170	10.9	5	conserved hypothet
750	2	T28000	171	10.9	5	hypothetical prote
751	2	T07380	171	10.9	5	invertase inhibito
752	2	E69711	171	10.9	5	stage III sporulat
753	2	H84709	171	10.9	5	probable glycine-r
754	2	S66014	172	10.9	5	single-stranded DN
755	2	H83542	172	10.9	5	hypothetical prote
756	2	F83585	172	10.9	5	probable sigma-70
757	2	A70864	172	10.9	5	hypothetical prote
758	2	S18035	173	10.9	5	pathogenesis-relat
759	2	S18034	173	10.9	5	pathogenesis-relat
760	2	S16524	173	10.9	5	thaumatin-like pro
761	2	C64161	173	10.9	5	hypothetical prote
762	2	G75554	173	10.9	5	hypothetical prote
763	2	F70109	173	10.9	5	hypothetical prote
764	2	A47303	173	10.9	5	FT2-F1 steroid rec
765	2	H95399	173	10.9	5	probable transcrip
766	2	C90006	173	10.9	5	hypothetical prote
767	2	S08535	173	10.9	5	hypothetical prote
768	2	S68274	174	10.9	5	myelin basic prote
769	2	E69038	174	10.9	5	inositol 1,4,5-tri
770	2	T13538	174	10.9	5	heat shock protein
771	2	F82786	174	10.9	5	conserved hypothet
772	2	E72262	175	10.9	5	conserved hypothet
773	2	AB1657	175	10.9	5	hypothetical prote
774	2	AG1285	175	10.9	5	hypothetical prote
775	2	I50016	175	10.9	5	MHC class I protei
776	2	JQ1232	176	10.9	5	calcineurin regula
777	2	T47036	176	10.9	5	hypothetical prote
778	2	S55970	176	10.9	5	hypothetical prote
779	2	AF0234	176	10.9	5	ribosomal protein
780	2	B70190	177	10.9	5	fimbrial protein [
781	2	S77911	177	10.9	5	conserved hypothet
782	2	B82221	177	10.9	5	hypothetical prote
783	2	AC2826	177	10.9	5	conserved hypothet
784	1	KIVZM1	178	10.9	5	conserved hypothet
785	2	A64115	178	10.9	5	thymidine kinase (
786	2	A35123	178	10.9	5	parB protein homol
787	2	B71824	178	10.9	5	ail protein precur
788	1	R5EG5	179	10.9	5	hypothetical protein
789	1	GZRT0	179	10.9	5	ribosomal protein
790	2	JC1221	179	10.9	5	secretory granule
791	2	H82921	179	10.9	5	calcineurin regula
792	2	B84466	179	10.9	5	conserved hypothet
793	2	B45613	180	10.9	5	40S ribosomal prot
794	2	T41339	180	10.9	5	surface antigen FU
795	2	T12914	180	10.9	5	hypothetical prote
796	2	C71869	180	10.9	5	hypothetical prote
797	2	S69721	181	10.9	5	hypothetical prote
798	2	JC7228	181	10.9	5	probable adenine p
799	2	AB0127	181	10.9	5	G-protein signalin
800	2	T05925	181	10.9	5	Syd protein [impor
801	2	I41321	181	10.9	5	hypothetical prote
802	1	KRBO2A	182	10.9	5	K99 fimbrial subun
803	2	T10180	182	10.9	5	keratin, 68K type
804	2	T04682	182	10.9	5	senescence-associa
805	2	B97604	182	10.9	5	senescence-associa
806	1	KNR2G2	183	10.9	5	hypothetical prote
807	2	T10192	183	10.9	5	glycine-rich cell
808	2	G97436	183	10.9	5	senescence-associa
809	2	AC2655	183	10.9	5	lemA protein (U661
810	2	T15618	183	10.9	5	conserved hypothet
811	2	AF0259	183	10.9	5	hypothetical prote
812	2	AD1815	184	10.9	5	probable phage-rel
813	2	AG0115	184	10.9	5	adenylate kinase [
814	2	C64592	184	10.9	5	probable membrane
815	2	A83316	184	10.9	5	hypothetical prote
816	2	H87676	184	10.9	5	hypothetical prote
817	2	D83234	184	10.9	5	hypothetical prote
818	2	A96615	185	10.9	5	probable dirigent
819	2	F69746	185	10.9	5	hypothetical prote
820	2	S14840	186	10.9	5	dihydrofolate redu
821	2	S76056	186	10.9	5	hypothetical prote
822	2	F96718	186	10.9	5	hypothetical prote
823	1	RDBE11	187	10.9	5	dihydrofolate redu
824	2	G70406	187	10.9	5	siroheme synthase
825	2	JH0240	187	10.9	5	aspartic proteinas
826	2	F81013	187	10.9	5	histidinol-phospha
827	2	B81957	187	10.9	5	probable phosphata
828	2	AH1870	187	10.9	5	hypothetical prote
829	2	G72748	187	10.9	5	hypothetical prote
830	2	C86895	188	10.9	5	conserved hypothet
831	2	A64332	188	10.9	5	probable phosphono
832	2	AC1579	188	10.9	5	hypothetical prote

#			len	n	Accession	Description
833	10.9	5	188	2	AI1225	hypothetical prote
834	10.9	5	188	2	B97875	conserved hypothet
835	10.9	5	189	2	A86699	thymidine kinase (
836	10.9	5	189	2	JC4072	virulence-associat
837	10.9	5	189	2	F82448	conserved hypothet
838	10.9	5	189	2	AD1224	precorrin decarbox
839	10.9	5	189	2	AF1577	precorrin decarbox
840	10.9	5	189	2	B95219	conserved hypothet
841	10.9	5	189	2	H98082	hypothetical prote
842	10.9	5	189	2	H70213	conserved hypothet
843	10.9	5	189	2	D69128	conserved hypothet
844	10.9	5	189	2	E97498	outer membrane ser
845	10.9	5	190	2	T06953	anthranilate synth
846	10.9	5	190	2	D75401	transcription repr
847	10.9	5	190	2	H59105	hypothetical prote
848	10.9	5	190	2	T26019	hypothetical prote
849	10.9	5	191	2	AB1768	thymidine kinase h
850	10.9	5	191	2	AH1392	thymidine kinase h
851	10.9	5	191	2	T19804	hypothetical prote
852	10.9	5	191	1	T19803	hypothetical prote
853	10.9	5	192	2	A64627	phosphoheptose iso
854	10.9	5	192	2	B71887	phosphoheptose iso
855	10.9	5	192	2	S70779	asparagine--tRNA l
856	10.9	5	192	2	G71089	hypothetical prote
857	10.9	5	193	1	B69043	conserved hypothet
858	10.9	5	193	2	D72474	probable deoxyribo
859	10.9	5	193	2	F64158	hypothetical prote
860	10.9	5	193	2	AF1981	hypothetical prote
861	10.9	5	193	2	T28200	hypothetical prote
862	10.9	5	193	2	AG2389	hypothetical prote
863	10.9	5	193	2	S39829	probable membrane
864	10.9	5	194	2	T18336	icmC protein - Leg
865	10.9	5	194	2	AD1754	proteinase (bacter
866	10.9	5	194	2	AI0241	hypothetical prote
867	10.9	5	195	2	S39777	hypothetical prote
868	10.9	5	195	2	T21997	actin beta - pig (
869	10.9	5	195	2	AD2782	hypothetical prote
870	10.9	5	195	2	A99346	hypothetical prote
871	10.9	5	195	1	D70571	hypothetical prote
872	10.9	5	196	2	AQJFNV	aequorin precursor
873	10.9	5	196	2	A30327	corticoliberin pre
874	10.9	5	196	2	E97561	hypothetical prote
875	10.9	5	196	2	E90273	conserved hypothet
876	10.9	5	196	2	G65039	hypothetical prote
877	10.9	5	196	2	B71232	hypothetical prote
878	10.9	5	197	2	F71248	hypothetical prote
879	10.9	5	197	2	H75212	hypothetical prote
880	10.9	5	197	2	T03442	glycine-rich prote
881	10.9	5	197	2	T44358	hypothetical prote
882	10.9	5	197	2	C64422	hypothetical prote
883	10.9	5	197	2	D97061	uncharacterized lo
884	10.9	5	197	2	AH1372	proteins involved
885	10.9	5	198	2	A57716	thiol-specific ant
886	10.9	5	198	2	T06813	dehydrin 3 garde
887	10.9	5	198	2	B97088	amidase, related t
888	10.9	5	198	2	AG3239	cell filamentation
889	10.9	5	199	2	B97987	thymidine kinase (
890	10.9	5	199	2	D95117	sigma-B activity i
891	10.9	5	199	2	D36131	hypothetical prote
892	10.9	5	199	2	T25260	conserved hypothet
893	10.9	5	199	2	B86655	probable periplasm
894	10.9	5	199	2	A81271	phosphoglycerate m
895	10.9	5	199	2	AD1230	weakly phosphoglyc
896	10.9	5	199	2	AG1583	hypothetical prote
897	10.9	5	200	2	E85047	hypothetical prote
898	10.9	5	200	2	C84283	hypothetical prote
899	10.9	5	200	2	B69168	probable esterasel
900	10.9	5	200	2	D95950	troponin I, fast s
901	10.9	5	201	2	A31484	hypothetical prote
902	10.9	5	201	1	E83780	brefeldin A resist
903	10.9	5	201	2	A45029	probable phage ant
904	10.9	5	202	2	G81897	1-acylglycerol-3-p
905	10.9	5	202	2	AG1280	1-acylglycerol-3-p
906	10.9	5	202	2	AG1643	1-acylglycerol-3-p
907	10.9	5	203	1	WMBPTB	gene 10 protein -
908	10.9	5	203	2	D81328	probable riboflavi
909	10.9	5	203	2	G81700	thymidylate kinase
910	10.9	5	203	2	JC7521	ribosomal protein
911	10.9	5	203	2	T02137	hypothetical prote
912	10.9	5	203	2	E64332	signal peptidase I
913	10.9	5	203	2	F90086	H repeat-associate
914	10.9	5	203	2	F90302	conserved hypothet
915	10.9	5	203	2	C85731	H repeat-associate
916	10.9	5	204	2	S35256	actin (clone Tac9)
917	10.9	5	204	2	T05677	hypothetical prote
918	10.9	5	205	2	T46553	recF protein [impo
919	10.9	5	205	2	A84717	hypothetical prote
920	10.9	5	205	2	B97302	transcription regu
921	10.9	5	205	2	A40305	transcription regu
922	10.9	5	206	2	A81231	50S ribosomal prot
923	10.9	5	206	2	D86832	uridine kinase (EC
924	10.9	5	206	2	T38724	probable peroxisom
925	10.9	5	206	2	AG2069	hypothetical prote
926	10.9	5	206	2	G86965	conserved hypothet
927	10.9	5	206	2	F64620	hypothetical prote
928	10.9	5	206	2	E71894	hypothetical prote
929	10.9	5	206	2	B28439	maturase-related h
930	10.9	5	207	2	T07381	glycine-rich prote
931	10.9	5	207	2	S67776	probable membrane
932	10.9	5	207	2	G70101	uridine kinase (ud
933	10.9	5	207	2	B89943	uridine kinase [im
934	10.9	5	207	2	D83187	hypothetical prote
935	10.9	5	207	2	AH1190	3-methyladenine DN
936	10.9	5	207	2	AH1548	3-methyladenine DN
937	10.9	5	207	2	C69780	hypothetical prote
938	10.9	5	207	2	G84983	5-amino-6-(5-phosp
939	10.9	5	208	2	T46896	merozoite surface
940	10.9	5	208	2	D88961	protein F59A7.6 [i
941	10.9	5	208	2	B86874	hypothetical prote
942	10.9	5	208	2	T09901	hypothetical prote
943	10.9	5	208	2	AB1485	hypothetical prote
944	10.9	5	208	2	AC1624	hypothetical prote
945	10.9	5	209	2	AI1261	Uridine kinase hom
946	10.9	5	209	2	AB1444	Uridine kinase hom
947	10.9	5	209	2	H82037	gp51 (Bacteriophag
948	10.9	5	210	2	S49577	conserved hypothet
949	10.9	5	210	2	E75553	chloramphenicol O-
950	10.9	5	210	2	E84121	uridine kinase - D
951	10.9	5	210	2	F69250	stage II sporulati
952	10.9	5	210	2	T40465	corrinoid methyltr
953	10.9	5	210	2	G69728	probable mitosis a
954	10.9	5	211	2	T36373	uridine kinase udk
955	10.9	5	211	2	S08522	hypothetical prote
956	10.9	5	211	2	T19859	probable sodium-de
957	10.9	5	211	2	T41218	hypothetical prote
958	10.9	5	211	2	AEI711	hypothetical prote
959	10.9	5	211	2	A60327	hypothetical prote
960	10.9	5	212	2	E81669	pilin precursor -
961	10.9	5	212	2	B82920	conserved hypothet
962	10.9	5	212	1	RDBEHS	hypothetical prote
963	10.9	5	213	2	A69310	dihydrofolate redu
964	10.9	5	213	2	A21177	proteasome, subuni
965	10.9	5	213	2	S68213	Ig light chain pre
966	10.9	5	213	2	A61043	Ig heavy chain (Ma
967	10.9	5	213	2	S63996	actin CA15 - sea s
968	10.9	5	214	2	S06788	pilin precursor -
969	10.9	5	214	2	S17680	fimbrial protein h
970	10.9	5	214	2	PC4202	fibrinolytic prote
971	10.9	5	214	2	C86967	monoclonal antibod
972	10.9	5	214	2	B96755	probable MADS box
973	10.9	5	214	2	F86197	probable membrane
974	10.9	5	214	2	A35063	hypothetical prote
975	10.9	5	214	2	S64108	hypothetical prote
976	10.9	5	215	2	T01349	hypothetical prote
977	10.9	5	215	2	F90312	hypothetical prote
978	10.9	5	215	2		conserved hypothet

979	5	10.9	215	2	AI1663	amino acid (glutam
980	5	10.9	216	2	S42716	calcineurin regula
981	5	10.9	216	1	E69128	ribosomal protein
982	5	10.9	216	2	A60331	pilin precursor -
983	5	10.9	216	2	A86633	conserved hypothet
984	5	10.9	216	2	T30657	hypothetical prote
985	5	10.9	216	2	B82612	hypothetical prote
986	5	10.9	216	2	AE2548	hypothetical prote
987	5	10.9	217	1	H71208	hypothetical prote
988	5	10.9	217	2	B71203	hypothetical prote
989	5	10.9	217	2	H70123	hypothetical prote
990	5	10.9	217	2	E95370	drought-induced pr
991	5	10.9	218	2	S37658	hypothetical prote
992	5	10.9	218	2	A64482	conserved hypothet
993	5	10.9	218	2	A69032	probable dedA prot
994	5	10.9	218	2	A70964	hypothetical prote
995	5	10.9	218	2	T23318	mutants block spor
996	5	10.9	218	2	C69712	hypothetical prote
997	5	10.9	218	2	B72470	d-ribulose-5-phosp
998	5	10.9	219	2	A82888	uncharacterized se
999	5	10.9	219	2	E97183	Ig heavy chain (Ma
1000	5	10.9	220	2	S68211	

ALIGNMENTS

RESULT 1
D70128
conserved hypothetical protein BB0228 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: D70128
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: D70128
A;Status: preliminary: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-971 <KLE>
A;Cross-references: GB:AE001133; GB:AE000783; NID:g2688120; PIDN:AAC66621.1; PID:g268812
A;Experimental source: strain B31

Query Match   19.6%; Score 9; DB 2; Length 971;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches   9; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   4 TLSLKNDFK 12
        |||||||||
Db 653 TLSLKNDFK 661

RESULT 2
G90517
hypothetical protein MYPU_0470 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90517
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21261165; PMID:11353084
A;Accession: G90517
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-88 <KUR>
A;Cross-references: GB:AL445566; PID:g14089460; PIDN:CAC13220.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:

A;Gene: MYPU_0470
A;Genetic code: SGC3

Query Match   15.2%; Score 7; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3 KTLSLKN 9
        |||||||
Db   20 KTLSLKN 26

RESULT 3
S46737
hypothetical protein YHR038w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein H8179.10
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C;Accession: S46737
R;Du, Z.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 8179.
A;Reference number: S46737
A;Accession: S46737
A;Molecule type: DNA
A;Residues: 1-230 <DUZ>
A;Cross-references: EMBL:U00062; NID:g488162; PID:g488171; GSPDB:GN00008; MIPS:YHR038
C;Genetics:
A;Gene: MIPS:YHR038w
A;Map position: 8R

Query Match   15.2%; Score 7; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 13;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   34 RLFNRSF 40
        |||||||
Db   17 RLFNRSF 23

RESULT 4
F83488
probable ATP-binding component of ABC transporter PA1256 [imported] - Pseudomonas aer
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83488
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: F83488
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <STO>
A;Cross-references: GB:AE004555; GB:AE004091; NID:g9947183; PIDN:AAG04645.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1256

Query Match   15.2%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 13;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   22 IIGGSGS 28
        |||||||
Db   32 IIGGSGS 38

RESULT 5
H95873
probable amino acid ABC transporter ATP-binding protein SMb20266 [imported] - Sinorhi

C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: H95873
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
ster, E.W.
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: H95873
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC48656.1; PID:g15140128; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMb20266
A;Genome: plasmid

Query Match   15.2%;   Score 7;   DB 2;   Length 240;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   22   IIGGSGS   28
          |||||||
Db   32   IIGGSGS   38

RESULT 6
G98155
probable ATP-binding component of ABC transporter PA1256 [imported] - Agrobacterium tume
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: G98155
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: G98155
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88769.1; PID:g15158515; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_402
A;Map position: linear chromosome

Query Match   15.2%;   Score 7;   DB 2;   Length 240;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   22   IIGGSGS   28
          |||||||
Db   32   IIGGSGS   38

RESULT 7
AE3132
hypothetical protein Atu4681 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AE3132
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45475.1; PID:g17743181; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4681
A;Map position: linear chromosome

Query Match   15.2%;   Score 7;   DB 2;   Length 240;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   22   IIGGSGS   28
          |||||||
Db   32   IIGGSGS   38

RESULT 8
B70483
ABC transporter - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Feb-2001
C;Accession: B70483
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: B70483
A;Status: preliminary: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-253 <AQF>
A;Cross-references: GB:AE000774; NID:g2984324; PIDN:AAC07853.1; PID:g2984332; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: abcT12
C;Superfamily: inner membrane protein malK: ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;20-218/Domain: ATP-binding cassette homology <ABC>
F;37-44/Region: nucleotide-binding motif A (P-loop)

Query Match   15.2%;   Score 7;   DB 2;   Length 253;
Best Local Similarity   100.0%;   Pred. No. 14;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   22   IIGGSGS   28
          |||||||
Db   35   IIGGSGS   41

RESULT 9
S70367
stem cell factor short form precursor - quail
C;Species: Coturnix coturnix (quail)
C;Date: 06-Dec-1996 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
C;Accession: S70367
R;Petitte, J.N.; Kulik, M.J.
Biochim. Biophys. Acta 1307, 149-151, 1996
A;Title: Cloning and characterization of cDNAs encoding two forms of avian stem cell
A;Reference number: S70366; MUID:96283808
A;Accession: S70367
A;Molecule type: mRNA
A;Residues: 1-253 <PET>
A;Cross-references: EMBL:U43079; NID:g1150877; PIDN:AAC59934.1; PID:g1150878
C;Superfamily: mouse mast cell growth factor
C;Keywords: growth factor; transmembrane protein

F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-253/Product: stem cell factor short form #status predicted <MAT>
F;192-216/Domain: transmembrane #status predicted <TMM>

Query Match   15.2%; Score 7; DB 2; Length 253;
Best Local Similarity   100.0%; Pred. No. 14;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   32 FFRLFNR 38
          |||||||
Db   144 FFRLFNR 150

RESULT 10
S70366
stem cell factor long form precursor - quail
C;Species: Coturnix coturnix (quail)
C;Date: 06-Dec-1996 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
C;Accession: S70366
R;Petitte, J.N.; Kulik, M.J.
Biochim. Biophys. Acta 1307, 149-151, 1996
A;Title: Cloning and characterization of cDNAs encoding two forms of avian stem cell fac
A;Reference number: S70366; MUID:96283808
A;Accession: S70366
A;Molecule type: mRNA
A;Residues: 1-287 <PET>
A;Cross-references: EMBL:U43078; NID:g1150875; PIDN:AAC59933.1; PID:g1150876
C;Superfamily: mouse mast cell growth factor
C;Keywords: growth factor; transmembrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-287/Product: stem cell factor long form #status predicted <MAT>
F;226-250/Domain: transmembrane #status predicted <TMM>

Query Match   15.2%; Score 7; DB 2; Length 287;
Best Local Similarity   100.0%; Pred. No. 16;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   32 FFRLFNR 38
          |||||||
Db   144 FFRLFNR 150

RESULT 11
T48633
hypothetical protein T15N1.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48633
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T48633
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <BEV>
A;Cross-references: EMBL:AL163792
A;Experimental source: cultivar Columbia; BAC clone T15N1
C;Genetics:
A;Map position: 5
A;Introns: 150/2; 232/1; 307/1
A;Note: T15N1.90

Query Match   15.2%; Score 7; DB 2; Length 370;
Best Local Similarity   100.0%; Pred. No. 20;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   25 GSGSLST 31
          |||||||
Db   118 GSGSLST 124

RESULT 12
I55604

platelet glycoprotein Ib beta chain precursor, endothelial splice form - human
N;Alternate names: membrane glycoprotein Ib beta chain (GPIb)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 22-Jun-1999
C;Accession: I55604
R;Kelly, M.D.; Essex, D.W.; Shapiro, S.S.; Meloni, F.J.; Druck, T.; Huebner, K.; Konk
J. Clin. Invest. 93, 2417-2424, 1994
A;Title: Complementary DNA cloning of the alternatively expressed endothelial cell gl
A;Reference number: I55604; MUID:94259799
A;Accession: I55604
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-411 <KEL>
A;Cross-references: GB:L20860; NID:g438627; PIDN:AAA20398.1; PID:g517494
C;Genetics:
A;Gene: GDB:GP1BB
A;Cross-references: GDB:128731; OMIM:138720
A;Map position: 22q11.21-22q11.23
A;Introns: 209/1
C;Complex: heterodimer with platelet glycoprotein Ib alpha chain (PIR:NBHUIA)
C;Superfamily: platelet glycoprotein Ib beta chain; leucine-rich alpha-2-glycoprotein
C;Keywords: alternative splicing; blood coagulation; duplication; glycoprotein; heter
F;224-249/Domain: proteoglycan amino-terminal homology <PAH>
F;261-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>
F;292-338/Domain: proteoglycan carboxyl-terminal homology <PCH>
F;353-377/Domain: transmembrane #status predicted <TRM>
F;271/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;396/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match   15.2%; Score 7; DB 1; Length 411;
Best Local Similarity   100.0%; Pred. No. 21;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   24 GGGSLS 30
          ||||||
Db   45 GGGSLS 51

RESULT 13
AC1112
PTS beta-glucoside-specific enzyme IIC component homolog lmo0298 [imported] - Listeri
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 31-Dec-2001
C;Accession: AC1112
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1112
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00825.1; PID:g16409663; GSPDB:GN0177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0298
C;Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-dep

Query Match   15.2%; Score 7; DB 2; Length 437;
Best Local Similarity   100.0%; Pred. No. 23;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   22 IIGGSGS 28
          |||||||
Db   293 IIGGSGS 299

RESULT 14
AG1473

PTS beta-glucoside-specific enzyme IIC component homolog lin0326 [imported] - Listeria i
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 31-Dec-2001
C;Accession: AG1473
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1473
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95559.1; PID:g16412755; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0326
C;Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-depend

Query Match   15.2%; Score 7; DB 2; Length 437;
Best Local Similarity   100.0%; Pred. No. 23;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   22 IIGGSGS 28
          |||||||
Db   293 IIGGSGS 299

RESULT 15
AD2284
hypothetical protein alr3827 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AD2284
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2284
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75526.1; PID:g17132961; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3827
C;Superfamily: Synechocystis hypothetical protein sll0827

Query Match   15.2%; Score 7; DB 2; Length 472;
Best Local Similarity   100.0%; Pred. No. 24;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   39 SFTQALG 45
          |||||||
Db   408 SFTQALG 414

Search completed: November 5, 2002, 11:06:57
Job time : 69.8507 secs

```
           GenCore version 5.1.3
     Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   November  5, 2002, 10:58:02 ; Search time 12.3582 seconds
                               (without alignments)
                               144.123 Million cell updates/sec

Title:          US-09-833-017B-2
Perfect score:  46
Sequence:       1 MKKTLSLKNDFKEIKTDELE........GSLSTFFRLFNRSFTQALGK 46

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       105224 seqs, 38719550 residues

Word size :     0

Total number of hits satisfying chosen parameters:   105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :      SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match Score	Length	DB	ID	Description
1	9	19.6	971	1	Y228_BORBU	O51246 borrelia bu
2	7	15.2	230	1	FIL1_YEAST	P38771 saccharomyc
3	7	15.2	500	1	ABF1_TRIRE	Q92455 trichoderma
4	7	15.2	500	1	XYL1_TRIKO	P48792 trichoderma
5	7	15.2	558	1	RELB_MOUSE	Q04863 mus musculu
6	7	15.2	579	1	RELB_HUMAN	Q01201 homo sapien
7	7	15.2	591	1	YN48_YEAST	P42846 saccharomyc
8	7	15.2	788	1	UGS3_SOLTU	Q43847 solanum tub
9	7	15.2	791	1	K6PP_RABIT	P47859 oryctolagus
10	6	13.0	88	1	PSD8_DICDI	P02889 dictyosteli
11	6	13.0	110	1	CYC6_PORPU	P51200 porphyra pu
12	6	13.0	114	1	HV3B_HUMAN	P01763 homo sapien
13	6	13.0	122	1	HV3A_HUMAN	P01762 homo sapien
14	6	13.0	142	1	RADC_COXBU	O85403 coxiella bu
15	6	13.0	155	1	Y805_AQUAE	O66989 aquifex aeo
16	6	13.0	169	1	CCAC_CAVPO	O35505 cavia porce
17	6	13.0	177	1	CCAC_CHICK	O73707 gallus gall
18	6	13.0	214	1	Y540_AQUAE	O66819 aquifex aeo
19	6	13.0	216	1	CAT1_STAAU	P00485 staphylococ
20	6	13.0	219	1	PSB9_HUMAN	P28065 homo sapien
21	6	13.0	219	1	PSB9_MOUSE	P28076 mus musculu
22	6	13.0	219	1	PSB9_MUSDU	O35521 mus dunni (
23	6	13.0	219	1	PSB9_MUSPL	O35523 mus platyth
24	6	13.0	219	1	PSB9_MUSSI	O35524 mus spicile
25	6	13.0	219	1	PSB9_RAT	P28077 rattus norv
26	6	13.0	226	1	PSB6_SCHPO	O43063 schizosacch
27	6	13.0	234	1	PSB6_TOBAC	P93395 nicotiana t
28	6	13.0	237	1	PSB6_RAT	P28073 rattus norv
29	6	13.0	237	1	Y417_METJA	Q57860 methanococc
30	6	13.0	238	1	PSB6_MOUSE	Q60692 mus musculu
31	6	13.0	239	1	PSB6_HUMAN	P28072 homo sapien
32	6	13.0	257	1	Y125_PYRHO	O57865 pyrococcus
33	6	13.0	257	1	YM64_PYRAB	Q9V2f1 pyrococcus
34	6	13.0	273	1	DAPF_HELPJ	Q9zlr5 helicobacte
35	6	13.0	273	1	DAPF_HELPY	O25290 helicobacte
36	6	13.0	276	1	OCCT_AGRTU	P35121 agrobacteri
37	6	13.0	281	1	CCAD_MOUSE	Q99246 mus musculu
38	6	13.0	285	1	CCAS_CHICK	O42398 gallus gall
39	6	13.0	287	1	MTAP_CAEEL	Q09438 caenorhabdi
40	6	13.0	296	1	SCF_CHICK	Q09108 gallus gall
41	6	13.0	297	1	CEBB_MOUSE	P28033 mus musculu
42	6	13.0	309	1	CEBB_RAT	P21272 rattus norv
43	6	13.0	311	1	HLGA_STAAM	P31714 staphylococ
44	6	13.0	314	1	OPPC_HAEIN	P45053 haemophilus
45	6	13.0	316	1	FMT_BUCAI	P57564 buchnera ap
46	6	13.0	317	1	Y441_CHLPN	Q9z8a2 chlamydia p
47	6	13.0	329	1	FATC_VIBAN	P37737 vibrio angu
48	6	13.0	342	1	ISPB_HAEIN	P44916 haemophilus
49	6	13.0	342	1	FLGI_HELPJ	Q9zmi9 helicobacte
50	6	13.0	344	1	FLGI_HELPY	O25028 helicobacte
51	6	13.0	345	1	ALF_MYCTU	O06313 mycobacteri
52	6	13.0	345	1	CATD_PIG	P00795 sus scrofa
53	6	13.0	348	1	CEBB_HUMAN	P17676 homo sapien
54	6	13.0	349	1	CEBB_BOVIN	O02755 bos taurus
55	6	13.0	354	1	YGA1_YEAST	P53199 s putative
56	6	13.0	365	1	FXF1_HUMAN	Q12946 homo sapien
57	6	13.0	372	1	CATD_SHEEP	Q9mzs8 ovis aries
58	6	13.0	373	1	INVE_SALTY	P35671 salmonella
59	6	13.0	375	1	ACT6_DIPDE	P53459 diphyllobot
60	6	13.0	387	1	T1SX_MYCPN	P75159 mycoplasma
61	6	13.0	390	1	ASPP_AEDAE	Q03168 aedes aeyp
62	6	13.0	393	1	CATD_BOVIN	P80209 bos taurus
63	6	13.0	393	1	CYB_VENIN	O48334 venturia in
64	6	13.0	396	1	YB2J_CAEEL	P91119 caenorhabdi
65	6	13.0	405	1	CARF_NEUCR	Q01294 neurospora
66	6	13.0	407	1	WCAC_ECOLI	P71237 escherichia
67	6	13.0	410	1	CATD_RAT	P24268 rattus norv
68	6	13.0	412	1	CATD_MOUSE	P18242 mus musculu
69	6	13.0	414	1	CATD_HUMAN	P07339 homo sapien
70	6	13.0	417	1	PYRC_THEAC	Q9him0 thermoplasm
71	6	13.0	442	1	DSRA_ARCFU	O59109 archaeoglob
72	6	13.0	442	1	SYH_HELPJ	Q9zk27 helicobacte
73	6	13.0	479	1	SYH_HELPY	P56455 helicobacte
74	6	13.0	484	1	APRA_PSEAE	Q03023 pseudomonas
75	6	13.0	494	1	RIO1_YEAST	Q12196 saccharomyc
76	6	13.0	497	1	STN1_YEAST	P38960 saccharomyc
77	6	13.0	499	1	UXAA_BACSU	O34673 bacillus su
78	6	13.0	499	1	ABFB_ASPNG	P42255 aspergillus
79	6	13.0	506	1	NU4C_MARPO	P06263 marchantia
80	6	13.0	518	1	Y619_METJA	Q58036 methanococc
81	6	13.0	521	1	LEU1_BACSU	P94565 bacillus su
82	6	13.0	538	1	GAG_HV2D2	P15832 human immun
83	6	13.0	550	1	RO60_XENLA	P42700 xenopus lae
84	6	13.0	550	1	MPI0_XENLA	P30308 xenopus lae
85	6	13.0	551	1	MPI1_XENLA	P30309 xenopus lae
86	6	13.0	561	1	CH60_BUCAP	O59177 buchnera ap
87	6	13.0	569	1	YEHU_ECOLI	P33357 escherichia
88	6	13.0	572	1	CH63_HELVI	P25420 heliothis v
89	6	13.0	573	1	MPI3_XENLA	P30311 xenopus lae
90	6	13.0	573	1	CH60_CRIGR	P18687 cricetulus
91	6	13.0	573	1	CH60_HUMAN	P10809 homo sapien
92	6	13.0	599	1	CH60_MOUSE	P19226 mus musculu
93	6	13.0	617	1	MPI2_XENLA	P30310 xenopus lae
94	6	13.0	620	1	SYP_SYNY3	P73942 synechocyst
95	6	13.0	634	1	S6A6_HUMAN	P31641 homo sapien
96	6	13.0	635	1	SL56_RAT	O70247 rattus norv
97	6	13.0	636	1	SL56_HUMAN	Q9y289 homo sapien
98	6	13.0	637	1	SL56_HUMAN	Q9xt77 oryctolagus
99	6	13.0	639	1	MUTA_PROFR	P11652 propionibac
100	6	13.0	664	1	ACTA_LISMO	P33379 listeria mo
101	6	13.0	699	1	UVRB_THEMA	Q9x282 thermotoga
102	6	13.0	703	1	YQOA_CAEEL	Q09299 caenorhabdi
103	6	13.0	717	1	ML51_HUMAN	O15234 homo sapien
104	6	13.0	735	1	AMYM_BACST	P19531 bacillus st
105	6	13.0	759	1	CNG1_CHICK	Q90805 gallus gall
106	6	13.0	759	1	PMT6_YEAST	P42934 saccharomyc

No.	n	Score	Len	1	ID	Accession	Organism
107	6	13.0	830	1	GYRA_BUCAI	P57277	buchnera ap
108	6	13.0	833	1	VIRA_AGRT5	P18540	agrobacteri
109	6	13.0	841	1	YAGX_ECOLI	P77802	escherichia
110	6	13.0	879	1	YE07_MYCPN	P75377	mycoplasma
111	6	13.0	887	1	HMDH_CRIGR	P00347	cricetulus
112	6	13.0	887	1	HMDH_MESAU	P09610	mesocricetu
113	6	13.0	887	1	HMDH_RAT	P51639	rattus norv
114	6	13.0	888	1	HMDH_HUMAN	P04035	homo sapien
115	6	13.0	888	1	HMDH_RABIT	Q29512	oryctolagus
116	6	13.0	970	1	Y087_BUCAI	P57189	buchnera ap
117	6	13.0	1049	1	NMD2_SCHPO	O13824	schizosacch
118	6	13.0	1075	1	PST2_SCHPO	O13919	schizosacch
119	6	13.0	1126	1	RPOB_SULAC	P11513	sulfolobus
120	6	13.0	1150	1	APMU_PIG	P12021	sus scrofa
121	6	13.0	1373	1	RPOB_RICCN	Q9rh41	rickettsia
122	6	13.0	1373	1	RPOB_RICMA	Q9rh43	rickettsia
123	6	13.0	1374	1	RPOB_RICPR	O52271	rickettsia
124	6	13.0	1610	1	CCAD_MESAU	Q99244	mesocricetu
125	6	13.0	1787	1	UVRA_CHLMU	Q9pk60	chlamydia m
126	6	13.0	1807	1	ITB4_RAT	Q64632	rattus norv
127	6	13.0	1852	1	CCAS_CYPCA	P22316	cyprinus ca
128	6	13.0	1966	1	CCAF_HUMAN	O60840	homo sapien
129	6	13.0	2139	1	CCAC_MOUSE	Q01815	mus musculu
130	6	13.0	2161	1	CCAC_HUMAN	Q01668	homo sapien
131	6	13.0	2169	1	CCAC_RAT	P22002	rattus norv
132	6	13.0	2171	1	CCAC_RABIT	P15381	oryctolagus
133	6	13.0	2190	1	CCAD_CHICK	O73700	gallus gall
134	6	13.0	2203	1	CCAD_RAT	P27732	rattus norv
135	6	13.0	2221	1	CCAC_HUMAN	Q13936	homo sapien
136	6	13.0	2569	1	LMA3_MOUSE	O61789	mus musculu
137	6	13.0	2747	1	FAF_DROME	P55824	drosophila
138	6	10.9	3911	1	AKA9_HUMAN	Q99996	h a-kinase
139	5	10.9	26	1	YFHA_KLEPN	P21710	klebsiella
140	5	10.9	27	1	ACH4_MOUSE	O70174	mus musculu
141	5	10.9	33	1	ACT_DICVI	O24733	dictyocaulu
142	5	10.9	40	1	HPT_RABIT	P19007	oryctolagus
143	5	10.9	45	1	CYC6_PROHO	P81244	prochloroth
144	5	10.9	55	1	A70A_DROMA	O18666	drosophila
145	5	10.9	56	1	RK33_GUITH	O78487	guillardia
146	5	10.9	60	1	LANP_STAEP	P19578	staphylococ
147	5	10.9	62	1	Y773_ARCFU	O29485	archaeoglob
148	5	10.9	63	1	PER_DROIM	O03294	drosophila
149	5	10.9	63	1	PER_DROMS	Q04535	drosophila
150	5	10.9	65	1	PER_DROMO	O03295	drosophila
151	5	10.9	65	1	RK33_PORPU	P51255	porphyra pu
152	5	10.9	66	1	PER_DROSA	Q04538	drosophila
153	5	10.9	67	1	YF82_ARCFU	O28690	archaeoglob
154	5	10.9	69	1	MCBA_ECOLI	P05834	escherichia
155	5	10.9	70	1	Y52A_METJA	P19285	thermoprote
156	5	10.9	82	1	Y9KD_BPBF2	P81309	methanococc
157	5	10.9	83	1	PER_DRORO	P19195	bacteriopha
158	5	10.9	86	1	CYSD_CHLLT	O03296	guillardia
159	5	10.9	87	1	IM13_ARATH	P20958	chlorobium
160	5	10.9	87	1	RL27_TREPA	Q9xh48	arabidopsis
161	5	10.9	88	1	GP42_BPSP1	O87725	treponema p
162	5	10.9	88	1	PER_DROTE	O48398	bacteriopha
163	5	10.9	88	1	S108_MOUSE	O26287	drosophila
164	5	10.9	89	1	SLTB_BPH30	P27005	mus musculu
165	5	10.9	90	1	RR17_GUITH	P08027	bacteriopha
166	5	10.9	91	1	YE14_HAEIN	O46903	guillardia
167	5	10.9	91	1	CBIN_SALTY	P44186	haemophilus
168	5	10.9	93	1	RS19_HELPJ	Q05595	salmonella
169	5	10.9	93	1	RS19_HELPY	Q9zjr7	helicobacte
170	5	10.9	94	1	ESA6_MYCTU	P56026	helicobacte
171	5	10.9	94	1	VA15_VACCC	Q57165	mycobacteri
172	5	10.9	94	1	VA15_VARV	P20992	vaccinia vi
173	5	10.9	95	1	I13B_HUMAN	P33840	variola vir
174	5	10.9	96	1	HOLI_BPA18	Q9y515	homo sapien
175	5	10.9	96	1	HOLI_BPA50	O37975	bacteriopha
176	5	10.9	97	1	YAFN_ECOLI	O37977	bacteriopha
177	5	10.9	97	1	YC40_METJA	Q47156	escherichia
178	5	10.9	101	1	VATF_METMA	Q58637	methanococc
179	5	10.9		1		Q60185	methanosarc

No.	n	Score	Len	1	ID	Accession	Organism
180	5	10.9	102	1	YAES_YEAST	P39725	saccharomyc
181	5	10.9	103	1	CYT1_MOUSE	P35173	mus musculu
182	5	10.9	104	1	YGY9_YEAST	P53069	saccharomyc
183	5	10.9	106	1	Y116_ADEO2	P03287	human adeno
184	5	10.9	108	1	Y488_ARCFU	O29762	archaeoglob
185	5	10.9	108	1	YDBL_ECOLI	P76076	escherichia
186	5	10.9	109	1	YE89_MYCTU	P71770	mycobacteri
187	5	10.9	109	1	FEL2_FELCA	P30440	felis silve
188	5	10.9	109	1	PER_LOXAL	Q25221	loxocera al
189	5	10.9	109	1	PER_MUSDO	Q25435	musca domes
190	5	10.9	109	1	PER_RHACO	Q26231	rhagoletis
191	5	10.9	109	1	PER_SYRPI	Q26612	syritta pip
192	5	10.9	110	1	CSGC_ECOLI	P52107	escherichia
193	5	10.9	110	1	PER_DROER	Q26288	drosophila
194	5	10.9	110	1	VL1_FPVL	P06456	avian papil
195	5	10.9	113	1	Y959_CAMJE	P24375	campylobact
196	5	10.9	113	1	PER_DROOR	Q9pnx6	drosophila
197	5	10.9	114	1	GHRL_BOVIN	Q26289	bos taurus
198	5	10.9	116	1	RK19_CYACA	Q9bdj6	cyanidium c
199	5	10.9	116	1	GHRL_CANFA	Q9tm06	canis famil
200	5	10.9	117	1	GHRL_HUMAN	Q9bef8	homo sapien
201	5	10.9	117	1	GVPK_HALME	Q9ubu3	halobacteri
202	5	10.9	117	1	LYB_BACSU	O02236	bacillus su
203	5	10.9	117	1	SP21_PAEPO	P10773	paenibacill
204	5	10.9	118	1	CRYP_CRYPA	O32720	cryphonectr
205	5	10.9	118	1	GVK2_HALN1	P52753	halobacteri
206	5	10.9	119	1	RBFA_BUCAI	P33963	buchnera ap
207	5	10.9	120	1	YE55_PYRHO	P57457	pyrococcus
208	5	10.9	120	1	RL35_SCHPO	O59124	schizosacch
209	5	10.9	122	1	S109_BOVIN	O74904	bos taurus
210	5	10.9	122	1	SVS5_MOUSE	P28783	mus musculu
211	5	10.9	124	1	PER_HIRPI	P30933	hirtodrosop
212	5	10.9	125	1	ACTS_PLEWA	Q25109	pleurodeles
213	5	10.9	125	1	PER_DROAN	P10994	drosophila
214	5	10.9	126	1	YEQ1_YEAST	Q03293	saccharomyc
215	5	10.9	128	1	IF2G_RABIT	P40045	oryctolagus
216	5	10.9	129	1	YF9A_THEMA	P33887	thermotoga
217	5	10.9	130	1	FABP_BLOTA	P58011	blomia trop
218	5	10.9	130	1	GLOV_HYACE	O17284	hyalophora
219	5	10.9	131	1	FABP_LEPDS	P81048	lepidoglyph
220	5	10.9	131	1	PER_ZAPTU	Q9u5p1	zaprionus t
221	5	10.9	132	1	YN64_YEAST	Q27135	saccharomyc
222	5	10.9	134	1	RL14_ARCFU	P53825	archaeoglob
223	5	10.9	138	1	CENA_MOUSE	O28364	mus musculu
224	5	10.9	139	1	STP4_PIG	O35216	sus scrofa
225	5	10.9	139	1	A85A_MYCMR	Q09821	mycobacteri
226	5	10.9	140	1	R17B_ARATH	Q9kh57	arabidopsis
227	5	10.9	140	1	R17A_ARATH	Q9sj36	arabidopsis
228	5	10.9	140	1	R17D_ARATH	P49205	arabidopsis
229	5	10.9	141	1	Y337_MYCPN	Q9lz17	mycoplasma
230	5	10.9	141	1	YC49_ARCFU	P75297	archaeoglob
231	5	10.9	141	1	PER_DROSR	O29019	drosophila
232	5	10.9	142	1	Y794_METJA	Q04537	methanococc
233	5	10.9	143	1	YEF5_YEAST	Q58204	saccharomyc
234	5	10.9	143	1	ASP2_STRTR	P32616	streptococc
235	5	10.9	143	1	DUT_CLODI	O30851	clostridium
236	5	10.9	147	1	GSPG_XANCP	O30931	xanthomonas
237	5	10.9	147	1	PER_DROPI	P31734	drosophila
238	5	10.9	147	1	PLIA_TRIFL	Q25206	trimeresuru
239	5	10.9	147	1	PLIB_TRIFL	P21755	trimeresuru
240	5	10.9	148	1	Y521_ARCFU	P21756	archaeoglob
241	5	10.9	148	1	Y565_META	O29729	methanococc
242	5	10.9	151	1	YEAL_ECOLI	Q57985	escherichia
243	5	10.9	154	1	YORX_PYRWO	P76240	pyrococcus
244	5	10.9	154	1	SP2C_STRPU	Q20298	strongyloce
245	5	10.9	155	1	ACT4_LYTPI	P04111	lytechinus
246	5	10.9	156	1	HS12_ORYSA	Q25380	oryza sativ
247	5	10.9	158	1	Y359_RICPR	P31673	rickettsia
248	5	10.9	159	1	BCCP_PSEAE	Q92dg9	pseudomonas
249	5	10.9		1	ISPF_PASMU	P37799	pasteurella
250	5	10.9		1	ISPF_VIBCH	P57954	pasteurella
251	5	10.9		1	ISPF_VIBCH	Q9kuj1	vibrio chol
252	5	10.9		1	ISPF_ECOLI	P36663	escherichia

Table A (bottom-left block)

Idx	10.9	5	Code	1	Entry	1	Accession	Organism
253	10.9	5	RT2C_ACTPL	1	160	1	P15376	actinobacil
254	10.9	5	ACT_PINCO	1	161	1	P24902	pinus conto
255	10.9	5	Y311_RICPR	1	161	1	Q9zdI6	rickettsia
256	10.9	5	Y447_METJA	1	161	1	Q57889	methanococc
257	10.9	5	Y07E_BPT4	1	164	1	P39491	bacteriopha
258	10.9	5	YPOC_BACSU	1	164	1	P39789	bacillus su
259	10.9	5	K2C5_BOVIN	1	166	1	P04262	bos taurus
260	10.9	5	LSPA_CHLMU	1	167	1	Q9piy8	chlamydia m
261	10.9	5	NADM_METJA	1	167	1	Q57961	methanococc
262	10.9	5	TCTP_LUMRU	1	167	1	O18477	lumbricus r
263	10.9	5	CALB_HUMAN	1	169	1	P06705	homo sapien
264	10.9	5	CALB_MOUSE	1	169	1	Q63810	mus musculu
265	10.9	5	CCME_BRAJA	1	169	1	P45401	bradyrhizob
266	10.9	5	RST4_AVESA	1	169	1	P50698	avena sativ
267	10.9	5	S3AB_BACSU	1	171	1	Q01368	bacillus su
268	10.9	5	ACT3_LYTPI	1	172	1	Q25379	lytechinus
269	10.9	5	ACTM_LYTPI	1	172	1	Q25381	lytechinus
270	10.9	5	SSB_BACSU	1	172	1	P37455	bacillus su
271	10.9	5	VP19_TBSVB	1	172	1	P50626	tomato bush
272	10.9	5	VP19_TBSVT	1	172	1	P50628	tomato bush
273	10.9	5	MBP_CHICK	1	173	1	P15720	gallus gall
274	10.9	5	PR1A_HORVU	1	173	1	P32937	hordeum vul
275	10.9	5	PR1C_HORVU	1	173	1	P32938	hordeum vul
276	10.9	5	TLP_WHEAT	1	173	1	P27357	triticum ae
277	10.9	5	YFHC_HAEIN	1	173	1	P44931	haemophilus
278	10.9	5	GRPE_METTH	1	174	1	O27350	methanobact
279	10.9	5	CALC_RAT	1	175	1	P29468	rattus norv
280	10.9	5	RI6B_YEAST	1	175	1	P05739	saccharomyc
281	10.9	5	ATPD_BUCAP	1	177	1	O51875	buchnera ap
282	10.9	5	AIL_YEREN	1	178	1	P16454	yersinia en
283	10.9	5	CALC_MOUSE	1	178	1	Q63811	mus musculu
284	10.9	5	KITH_MYXVA	1	178	1	P28851	myxoma viru
285	10.9	5	YC96_HAEIN	1	178	1	Q57519	haemophilus
286	10.9	5	PGSG_RAT	1	179	1	P04917	rattus norv
287	10.9	5	RK5_EUGGR	1	179	1	P21510	euglena gra
288	10.9	5	ARF1_PLAFO	1	180	1	Q25761	plasmodium
289	10.9	5	ARF_PLAFA	1	180	1	Q94650	plasmodium
290	10.9	5	APT2_YEAST	1	181	1	P36973	saccharomyc
291	10.9	5	FANC_ECOLI	1	181	1	P18103	escherichia
292	10.9	5	Y65L_HORVU	1	182	1	O48609	hordeum vul
293	10.9	5	K2CA_BOVIN	1	182	1	P04263	bos taurus
294	10.9	5	DIN1_RAPSA	1	183	1	P27626	raphanus sa
295	10.9	5	GRP2_ORYSA	1	183	1	P29834	oryza sativ
296	10.9	5	DYR_HSV57	1	186	1	P27421	herpesvirus
297	10.9	5	DYR_HSVSA	1	187	1	P09503	herpesvirus
298	10.9	5	COME_METJA	1	188	1	P58416	methanococc
299	10.9	5	EFP_PASMU	1	188	1	P57811	pasteurella
300	10.9	5	TRPG_CYAPA	1	190	1	P48261	cyanophora
301	10.9	5	KIP1_HUMAN	1	191	1	Q99828	homo sapien
302	10.9	5	KIP1_MOUSE	1	191	1	Q920f4	mus musculu
303	10.9	5	KIP1_RAT	1	191	1	Q9r010	rattus norv
304	10.9	5	KITH_STRGC	1	191	1	P47848	streptococc
305	10.9	5	LPCA_HELPJ	1	192	1	Q9zkz1	helicobacte
306	10.9	5	LPCA_HELPY	1	192	1	O25528	helicobacte
307	10.9	5	MOBA_PYRHO	1	192	1	O58708	pyrococcus
308	10.9	5	DEOC_AERPE	1	193	1	Q9y948	aeropyrum p
309	10.9	5	YBF9_YEAST	1	193	1	P34224	saccharomyc
310	10.9	5	YHHF_HAEIN	1	193	1	P44869	haemophilus
311	10.9	5	KITH_BACNA	1	195	1	O52951	bacillus su
312	10.9	5	AEO2_AEQVI	1	196	1	P02592	aequorea vi
313	10.9	5	CRF_HUMAN	1	196	1	P06850	homo sapien
314	10.9	5	RDVA_CLOHI	1	197	1	Q9znj6	clostridium
315	10.9	5	Y979_METJA	1	197	1	O58389	methanococc
316	10.9	5	PDX2_MOUSE	1	198	1	Q61171	mus musculu
317	10.9	5	PDX2_RAT	1	198	1	P35704	rattus norv
318	10.9	5	RSBX_BACSU	1	199	1	P17906	bacillus su
319	10.9	5	TDX2_BRUMA	1	199	1	Q17172	brugia mala
320	10.9	5	TRI_PONLE	1	201	1	P05547	pontastacus
321	10.9	5	P25_SCHPO	1	202	1	P30831	schizosacch
322	10.9	5	KTHY_CHLMU	1	203	1	Q9pkk5	chlamydia m
323	10.9	5	VP10_BPPRD	1	203	1	P28732	bacteriopha
324	10.9	5	Y260_METJA	1	203	1	Q57708	methanococc
325	10.9	5	CAHB_MOUSE	1	204	1	O70354	mus musculu

Table B (top-right block)

Idx	1	10.9	5	Entry	1	Code	Accession	Organism
326	1	10.9	5	205	1	FADD_MOUSE	Q61160	mus muscu
327	1	10.9	5	205	1	URK_BORBU	Q59190	borrelia bu
328	1	10.9	5	207	1	3MGH_LISIN	Q92d89	listeria in
329	1	10.9	5	207	1	3MGH_LISMO	P58621	listeria mo
330	1	10.9	5	207	1	IM22_YEAST	Q12328	saccharomyc
331	1	10.9	5	207	1	RID2_BUCAI	P57534	buchnera ap
332	1	10.9	5	210	1	CAT4_MORMO	P50869	morganella
333	1	10.9	5	210	1	HFA5_HAEIN	P45990	haemophilus
334	1	10.9	5	211	1	URK_BACSU	O32033	bacillus su
335	1	10.9	5	211	1	HFA3_HAEIN	P45988	haemophilus
336	1	10.9	5	212	1	RB17_HUMAN	Q9h0t7	homo sapien
337	1	10.9	5	212	1	DYR_HSVSC	P22573	herpesvirus
338	1	10.9	5	213	1	HFA2_HAEIN	P14212	haemophilus
339	1	10.9	5	213	1	HFA4_HAEIN	P45989	haemophilus
340	1	10.9	5	213	1	PSMB_ARCFU	Q9p996	archaeoglob
341	1	10.9	5	213	1	ACH2_LONAC	P23605	lonomia ach
342	1	10.9	5	214	1	YL31_HALMA	P20571	haloarcula
343	1	10.9	5	214	1	YGK1_MYCLE	Q49642	mycobacteri
344	1	10.9	5	215	1	YGK1_YEAST	P53144	saccharomyc
345	1	10.9	5	216	1	HFA1_HAEIN	O03846	haemophilus
346	1	10.9	5	216	1	RS5_METTH	O26131	methanobact
347	1	10.9	5	218	1	S3AH_BACSU	P49785	bacillus su
348	1	10.9	5	218	1	YC35_METTH	O27303	methanobact
349	1	10.9	5	218	1	YE58_METJA	Q58853	methanococc
350	1	10.9	5	218	1	YQ37_MYCTU	P71936	mycobacteri
351	1	10.9	5	221	1	YD1E_SCHPO	Q10244	schizosacch
352	1	10.9	5	222	1	FTSQ_CORGL	P94336	corynebacte
353	1	10.9	5	222	1	RISB_SPIOL	Q9xh32	spinacia ol
354	1	10.9	5	223	1	COAT_TRVCA	P05070	tobacco rat
355	1	10.9	5	224	1	GSHU_CAEEL	Q95003	caenorhabdi
356	1	10.9	5	224	1	PSMB_METJA	O58634	methanococc
357	1	10.9	5	224	1	Y088_MYCTU	Q10885	mycobacteri
358	1	10.9	5	224	1	YXEN_BACSU	P54953	bacillus su
359	1	10.9	5	227	1	DJB8_MOUSE	Q9gyi7	mus musculu
360	1	10.9	5	228	1	PCRB_BACSU	O34790	bacillus su
361	1	10.9	5	229	1	GLC8_YEAST	P41818	saccharomyc
362	1	10.9	5	229	1	XYN1_TRIRE	P36218	trichoderma
363	1	10.9	5	233	1	DEOD_TREPA	O83716	treponema p
364	1	10.9	5	233	1	YSR3_CAEEL	Q09951	caenorhabdi
365	1	10.9	5	234	1	EMB2_HUMAN	P35709	cavia porce
366	1	10.9	5	236	1	ABME_HUMAN	P41238	homo sapien
367	1	10.9	5	236	1	Y545_RICPR	Q9zd05	rickettsia
368	1	10.9	5	237	1	CD63_MOUSE	P41731	mus musculu
369	1	10.9	5	237	1	CD63_RAT	P28648	rattus norv
370	1	10.9	5	237	1	LAPT_PASHA	P49618	pasteurella
371	1	10.9	5	238	1	YB92_YEAST	P38331	saccharomyc
372	1	10.9	5	238	1	YPE2_RHORU	P23139	rhodospiril
373	1	10.9	5	239	1	ARTI_HAEIN	P45091	haemophilus
374	1	10.9	5	239	1	RP35_BACTK	P26763	bacillus th
375	1	10.9	5	239	1	VIB1_AGRT9	P05350	agrobacteri
376	1	10.9	5	240	1	H12_VOLCA	Q08865	volvox cart
377	1	10.9	5	240	1	LEF5_GVCL	P41727	cryptophleb
378	1	10.9	5	240	1	YMFC_HAEIN	P44827	haemophilus
379	1	10.9	5	241	1	Y293_MYCPN	P75367	mycoplasma
380	1	10.9	5	243	1	GUB_BACLI	P27051	bacillus li
381	1	10.9	5	243	1	TONB_PSEPU	Q05613	pseudomonas
382	1	10.9	5	244	1	Y293_MYCGE	P47535	mycoplasma
383	1	10.9	5	246	1	VGLL_HSV7J	P52509	human herpe
384	1	10.9	5	247	1	IF6_ARATH	O22290	arabidopsis
385	1	10.9	5	247	1	RL7_ARATH	O42208	arabidopsis
386	1	10.9	5	248	1	GRL1_RAT	Q06605	rattus norv
387	1	10.9	5	248	1	NUKC_MAIZE	P66670	zea mays (m
388	1	10.9	5	250	1	MTRD_METMA	P80653	methanosarc
389	1	10.9	5	250	1	YCW6_YEAST	P25659	saccharomyc
390	1	10.9	5	251	1	TX20_HUMAN	Q9umr3	homo sapien
391	1	10.9	5	252	1	PMM_CANAL	P31353	candida alb
392	1	10.9	5	252	1	TRPA_AERPE	Q9y8t3	aeropyrum p
393	1	10.9	5	252	1	TRY1_DROME	P52905	drosophila
394	1	10.9	5	252	1	YC88_METJA	O58684	methanococc
395	1	10.9	5	252	1	YKC6_CAEEL	P41997	caenorhabdi
396	1	10.9	5	253	1	NIKD_ECOLI	P33593	escherichia
397	1	10.9	5	254	1	MURI_AQUAE	O66662	aquifex aeo
398	1	10.9	5	254	1	RACD_DICDI	P34150	dictyosteli

Left column block

No.		Len			Identifier		Organism
399	1	255	10.9	5	PPNK_RICPR	1	rickettsia
400	1	255	10.9	5	THIG_CLOAB	1	clostridium
401	1	256	10.9	5	THIG_ECO57	1	escherichia
402	1	256	10.9	5	THIG_ECOLI	1	escherichia
403	1	256	10.9	5	THIG_SALTY	1	salmonella
404	1	256	10.9	5	THIG_VIBCH	1	vibrio chol
405	1	256	10.9	5	YKH1_YEAST	1	saccharomyc
406	1	256	10.9	5	YP96_CAEEL	1	caenorhabdi
407	1	257	10.9	5	Y453_AQUAE	1	aquifex aeo
408	1	257	10.9	5	YCP3_SYNPY	1	synechococc
409	1	260	10.9	5	MTC1_CHVN1	1	chlorella v
410	1	261	10.9	5	PSB7_YEAST	1	saccharomyc
411	1	261	10.9	5	YA54_AQUAE	1	aquifex aeo
412	1	262	10.9	5	ASH2_MOUSE	1	mus musculu
413	1	262	10.9	5	MSA2_PLAFC	1	plasmodium
414	1	262	10.9	5	Y4JP_RHISN	1	rhizobium s
415	1	264	10.9	5	AG15_BRANA	1	brassica na
416	1	264	10.9	5	YDCV_ECOLI	1	escherichia
417	1	265	10.9	5	RL7A_FUGRU	1	fugu rubrip
418	1	265	10.9	5	RL7A_HUMAN	1	homo sapien
419	1	265	10.9	5	RL7A_MOUSE	1	mus musculu
420	1	265	10.9	5	YM40_MYCTU	1	mycobacteri
421	1	266	10.9	5	PTNC_ECOLI	1	escherichia
422	1	266	10.9	5	SP2G_CLOAB	1	clostridium
423	1	266	10.9	5	XKDC_BACSU	1	bacillus su
424	1	267	10.9	5	BLO5_PSEAE	1	pseudomonas
425	1	267	10.9	5	PSB7_SCHPO	1	schizosacch
426	1	269	10.9	5	AROE_NEICI	1	neisseria c
427	1	269	10.9	5	CCHL_YEAST	1	saccharomyc
428	1	269	10.9	5	IL1B_HUMAN	1	homo sapien
429	1	270	10.9	5	TRPA_BUCDN	1	buchnera ap
430	1	270	10.9	5	VA32_VARV	1	variola vir
431	1	273	10.9	5	DHQU_MOUSE	1	mus musculu
432	1	273	10.9	5	FRA1_MOUSE	1	mus musculu
433	1	273	10.9	5	DHQU_RAT	1	rattus norv
434	1	274	10.9	5	FMF4_ECOLI	1	escherichia
435	1	274	10.9	5	MCRA_ECOLI	1	escherichia
436	1	275	10.9	5	PSB7_HUMAN	1	homo sapien
437	1	276	10.9	5	PSB7_MOUSE	1	mus musculu
438	1	276	10.9	5	SLBP_DROME	1	drosophila
439	1	276	10.9	5	YRD4_CAEEL	1	caenorhabdi
440	1	276	10.9	5	FMF4_ECOLI	1	escherichia
441	1	277	10.9	5	MCRA_ECOLI	1	escherichia
442	1	277	10.9	5	PSB7_HUMAN	1	homo sapien
443	1	277	10.9	5	PSB7_MOUSE	1	mus musculu
444	1	277	10.9	5	PSB7_RAT	1	rattus norv
445	1	277	10.9	5	ACT_CALFI	1	calanus fin
446	1	278	10.9	5	Y110_MYCGE	1	mycoplasma
447	1	278	10.9	5	YE27_HAEIN	1	haemophilus
448	1	281	10.9	5	NAE1_THEMA	1	thermotoga
449	1	282	10.9	5	CDX4_MOUSE	1	mus musculu
450	1	282	10.9	5	ISPA_BUCAI	1	buchnera ap
451	1	283	10.9	5	FMKC_ECOLI	1	escherichia
452	1	283	10.9	5	INSK_ECOLI	1	escherichia
453	1	284	10.9	5	DAPF_BUCAI	1	buchnera ap
454	1	284	10.9	5	YNFH_ECOLI	1	escherichia
455	1	285	10.9	5	FMKB_ECOLI	1	escherichia
456	1	285	10.9	5	FMKD_ECOLI	1	escherichia
457	1	285	10.9	5	GLO4_YEAST	1	saccharomyc
458	1	285	10.9	5	Y539_AQUAE	1	aquifex aeo
459	1	287	10.9	5	CFA1_MYCTU	1	mycobacteri
460	1	287	10.9	5	TSX_SALTY	1	salmonella
461	1	289	10.9	5	ATPG_HAEIN	1	haemophilus
462	1	289	10.9	5	LYTB_AQUAE	1	aquifex aeo
463	1	290	10.9	5	EFTS_TREPA	1	treponema p
464	1	290	10.9	5	LEP_HELPJ	1	helicobacte
465	1	290	10.9	5	LEP_HELPY	1	helicobacte
466	1	290	10.9	5	YA57_METJA	1	methanococc
467	1	291	10.9	5	C552_PSEST	1	pseudomonas
468	1	291	10.9	5	IBP3_MOUSE	1	mus musculu
469	1	291	10.9	5	SUD1_ARCFU	1	archaeoglob
470	1	292	10.9	5	LICB_HAEIN	1	haemophilus
471	1	294	10.9	5	TSX_ECOLI	1	escherichia

Right column block

No.		Len			Identifier		Accession	Organism
472	1	294	10.9	5	TSX_ENTAE	1	P40785	enterobacte
473	1	294	10.9	5	TSX_KLEPN	1	P40786	klebsiella
474	1	296	10.9	5	CYOA_BUCAI	1	P57544	buchnera ap
475	1	296	10.9	5	MY88_MOUSE	1	P22366	mus musculu
476	1	298	10.9	5	END4_BACHD	1	Q9kd33	bacillus ha
477	1	298	10.9	5	RT03_ACACA	1	P46754	acanthamoeb
478	1	299	10.9	5	RT03_STRGC	1	O24756	streptococc
479	1	299	10.9	5	ERA_STRPN	1	Q9xdg9	streptococc
480	1	299	10.9	5	ISPE_DEIRA	1	Q9rr89	deinococcus
481	1	300	10.9	5	PHAG_PSEAE	1	Q51553	pseudomonas
482	1	300	10.9	5	VA32_VACCV	1	P21055	vaccinia vi
483	1	300	10.9	5	YHC1_HAEIN	1	P44541	haemophilus
484	1	301	10.9	5	SC14_KLULA	1	P24859	kluyveromyc
485	1	301	10.9	5	YFE4_YEAST	1	P43558	saccharomyc
486	1	302	10.9	5	OPPC_ECOLI	1	P77664	escherichia
487	1	302	10.9	5	OPPC_SALTY	1	P08006	salmonella
488	1	302	10.9	5	YCAN_ECOLI	1	P55836	escherichia
489	1	305	10.9	5	PYRB_BACHD	1	Q9k9v6	bacillus ha
490	1	305	10.9	5	TNE2_HUMAN	1	O95859	homo sapien
491	1	305	10.9	5	Y758_ARCFU	1	O29500	archaeoglob
492	1	305	10.9	5	YEDI_ECOLI	1	P46125	escherichia
493	1	306	10.9	5	IF2A_SCHPO	1	P56286	schizosacch
494	1	306	10.9	5	KI28_YEAST	1	P06242	saccharomyc
495	1	307	10.9	5	CC36_CAEEL	1	P34803	caenorhabdi
496	1	307	10.9	5	CIW8_MOUSE	1	Q9z2t1	mus musculu
497	1	307	10.9	5	HEM3_CAMJE	1	Q9phw9	campylobact
498	1	307	10.9	5	TRXB_MYCPU	1	Q98pk9	mycoplasma
499	1	307	10.9	5	YAC2_SCHPO	1	Q09816	schizosacch
500	1	308	10.9	5	VG10_BPBO3	1	Q37891	bacteriopha
501	1	308	10.9	5	Y325_METJA	1	Q57771	methanococc
502	1	309	10.9	5	ACT3_ECHGR	1	Q03342	echinococcu
503	1	309	10.9	5	DDL_PASMU	1	P57819	pasteurella
504	1	309	10.9	5	YESP_BACSU	1	O31519	bacillus su
505	1	311	10.9	5	NODI_RHILV	1	P08720	rhizobium l
506	1	311	10.9	5	RBF1_MOUSE	1	P22560	mus musculu
507	1	312	10.9	5	RTO2_ACACA	1	P46753	acanthamoeb
508	1	312	10.9	5	VP12_RDVA	1	Q05054	rice dwarf
509	1	312	10.9	5	VP12_RDVF	1	Q85443	rice dwarf
510	1	312	10.9	5	VP12_RDVO	1	Q86780	rice dwarf
511	1	313	10.9	5	PYRD_BACCL	1	P46539	bacillus ca
512	1	314	10.9	5	NODI_RHILP	1	P23718	rhizobium l
513	1	314	10.9	5	SIX3_CHICK	1	O42406	gallus gall
514	1	315	10.9	5	SINR_SALTY	1	P37459	salmonella
515	1	315	10.9	5	SX22_HUMAN	1	O15370	homo sapien
516	1	316	10.9	5	LORI_HUMAN	1	P23490	homo sapien
517	1	317	10.9	5	CGEB_BACSU	1	P42090	bacillus su
518	1	317	10.9	5	FEOA_BACSU	1	P40409	bacillus su
519	1	317	10.9	5	YZ07_AQUAE	1	O66402	aquifex aeo
520	1	318	10.9	5	YZ11_AQUAE	1	O66405	aquifex aeo
521	1	318	10.9	5	YZ20_AQUAE	1	O66411	aquifex aeo
522	1	318	10.9	5	YZ34_AQUAE	1	O66423	aquifex aeo
523	1	319	10.9	5	VMP_CERV	1	P05396	carnation e
524	1	320	10.9	5	Y678_METJA	1	Q58091	methanococc
525	1	320	10.9	5	YDJE_BACSU	1	O34768	bacillus su
526	1	320	10.9	5	YGI1_YEAST	1	P53156	saccharomyc
527	1	322	10.9	5	ACCA_ANTSP	1	P46316	antithamnio
528	1	322	10.9	5	ACT_PROCL	1	P45521	procambarus
529	1	323	10.9	5	CITG_HAEDU	1	O30827	haemophilus
530	1	323	10.9	5	RFC4_YEAST	1	P40339	saccharomyc
531	1	324	10.9	5	GC1_MOUSE	1	P01868	mus musculu
532	1	324	10.9	5	HME2_MOUSE	1	P09066	mus musculu
533	1	324	10.9	5	Y640_METJA	1	Q58057	methanococc
534	1	324	10.9	5	YF02_METJA	1	Q58897	methanococc
535	1	324	10.9	5	YIAU_ECOLI	1	P37682	escherichia
536	1	324	10.9	5	YPDA_BACSU	1	P50736	bacillus su
537	1	325	10.9	5	FRZB_BOVIN	1	Q95117	bos taurus
538	1	325	10.9	5	FRZB_HUMAN	1	Q92765	homo sapien
539	1	325	10.9	5	UCP5_HUMAN	1	O95258	homo sapien
540	1	325	10.9	5	UCP5_MOUSE	1	Q9z2b2	mus musculu
541	1	325	10.9	5	YQZ1_CAEEL	1	Q09319	caenorhabdi
542	1	326	10.9	5	MTC2_CHVP1	1	P31118	paramecium
543	1	326	10.9	5	OMPH_PHOPR	1	P29739	photobacter
544	1	327	10.9	5	ACT3_ARTSX	1	P18602	artemia sp.

No.	Score	n	ID		Len	Accession	Organism
545	10.9	5	GRE3_YEAST	1	327	P38715	saccharomyc
546	10.9	5	CAHB_HUMAN	1	328	O75493	homo sapien
547	10.9	5	CAHB_SHEEP	1	328	Q95203	ovis aries
548	10.9	5	YCDU_ECOLI	1	328	P75910	escherichia
549	10.9	5	GCAA_MOUSE	1	330	P01863	mus musculu
550	10.9	5	YDEZ_ECOLI	1	330	P77651	escherichia
551	10.9	5	ACT_COSCS	1	331	P30161	costaria co
552	10.9	5	AP1_HUMAN	1	331	P05412	homo sapien
553	10.9	5	AP1_PIG	1	331	P56432	sus scrofa
554	10.9	5	PLIB_AGKBL	1	331	P93233	agkistrodon
555	10.9	5	ACT1_SOLTU	1	332	P19622	homo sapien
556	10.9	5	HME2_HUMAN	1	332	P57216	buchnera ap
557	10.9	5	RNFD_BUCAI	1	332	P40400	bacillus su
558	10.9	5	SSUA_BACSU	1	332	P53228	saccharomyc
559	10.9	5	TAL2_YEAST	1	333	Q54480	synechocyst
560	10.9	5	YZ37_SYNY3	1	333	P01864	mus musculu
561	10.9	5	GCAB_MOUSE	1	335	P55150	anabaena fl
562	10.9	5	GVPN_ANAFL	1	335	Q96482	lycopersico
563	10.9	5	ACT1_LYCES	1	336	Q96483	lycopersico
564	10.9	5	ACT2_LYCES	1	336	P93586	solanum tub
565	10.9	5	ACT2_SOLTU	1	336	P93374	nicotiana t
566	10.9	5	ACT2_TOBAC	1	336	Q96484	lycopersico
567	10.9	5	ACT3_LYCES	1	336	Q96481	lycopersico
568	10.9	5	ACT4_TOBAC	1	336	P93372	nicotiana t
569	10.9	5	ACT5_SOLTU	1	336	P81228	solanum tub
570	10.9	5	ACT5_TOBAC	1	336	P93371	nicotiana t
571	10.9	5	ACT6_TOBAC	1	336	P93375	nicotiana t
572	10.9	5	ACT7_TOBAC	1	336	P81229	nicotiana t
573	10.9	5	ACT8_SOLTU	1	336	P93584	solanum tub
574	10.9	5	ACT9_SOLTU	1	336	Q9z3r6	solanum tub
575	10.9	5	AGLF_RHIME	1	336	P01866	rhizobium m
576	10.9	5	GCB_MOUSE	1	336	P39242	mus musculu
577	10.9	5	Y05G_BPT4	1	336	P58248	bacteriopha
578	10.9	5	A85A_MYCUL	1	337	P93585	mycobacteri
579	10.9	5	ACT4_MYCUL	1	337	Q9ka01	mycobacteri
580	10.9	5	PLSX_BACHD	1	337	Q61038	bacillus ha
581	10.9	5	TDA8_MOUSE	1	337	P29888	mus musculu
582	10.9	5	VF11_VACCP	1	337	O34736	vaccinia vi
583	10.9	5	YJMC_BACSU	1	337	O00154	bacillus su
584	10.9	5	CTE2_HUMAN	1	338	Q64559	homo sapien
585	10.9	5	CTE2_RAT	1	338	Q51464	rattus norv
586	10.9	5	FLIG_PSEAE	1	338	P38000	pseudomonas
587	10.9	5	YBD1_YEAST	1	338	Q09862	saccharomyc
588	10.9	5	YAF9_SCHPO	1	339	Q9zem7	schizosacch
589	10.9	5	ALF_STRGB	1	340	P17118	streptomyce
590	10.9	5	THPS_SULAC	1	340	O31284	sulfolobus
591	10.9	5	BMPB_BORAF	1	340	P17768	borrelia af
592	10.9	5	VP3_GFLV	1	341	P75302	grapevine f
593	10.9	5	Y33B_MYCPN	1	341	P39328	mycoplasma
594	10.9	5	YTFT_ECOLI	1	341	Q9xff4	escherichia
595	10.9	5	COBT_METSQ	1	342	P37253	methylobact
596	10.9	5	ILVC_BACSU	1	342	O14003	bacillus su
597	10.9	5	RFC3_SCHPO	1	342	Q58165	schizosacch
598	10.9	5	Y755_METJA	1	342	P19537	methanococc
599	10.9	5	EFB1_RAT	1	342	Q9x8r6	rattus norv
600	10.9	5	YIB9_YEAST	1	343	O06703	streptomyce
601	10.9	5	FOS_CYPCA	1	343	Q18491	bordetella
602	10.9	5	HPT_MUSSA	1	343	O50146	caenorhabdi
603	10.9	5	HPT_RAT	1	344	Q58030	thermus aqu
604	10.9	5	Y4KE_RHISN	1	344	P11290	methanococc
605	10.9	5	AROH_BUCAI	1	344	O69600	escherichia
606	10.9	5	AROH_BUCAP	1	344	P52796	mycobacteri
607	10.9	5	ALF_MYCLE	1	345	P40546	rattus norv
608	10.9	5	ALF_CORGL	1	345	P79702	saccharomyc
609	10.9	5	ALF_STRCO	1	345	Q62558	cyprinus ca
610	10.9	5	BBUR_BORBR	1	346	P06866	mus saxicol
611	10.9	5	AR20_CAEEL	1	347	P55525	rattus norv
612	10.9	5	ARC2_THETH	1	347	P57224	rhizobium s
613	10.9	5	Y613_METJA	1	347	P46245	buchnera ap
614	10.9	5	YIBD_ECOLI	1	347	O54459	buchnera ap
615	10.9	5	AROH_BUCAP	1	348	Q9pmj8	erwinia her
616	10.9	5	AROH_ERWHE	1	348		
617	10.9	5	FLGI_CAMJE	1	348		campylobact
618	10.9	5	VMAT_P1HA	1	348	P36355	human parai
619	10.9	5	VMAT_P1HC	1	348	Q01427	human parai
620	10.9	5	VMAT_SEND5	1	348	P27567	sendai viru
621	10.9	5	VMAT_SENDF	1	348	P17748	sendai viru
622	10.9	5	VMAT_SENDH	1	348	P03426	sendai viru
623	10.9	5	VMAT_SENDZ	1	348	P06446	sendai viru
624	10.9	5	GPDA_DROAE	1	349	Q27556	drosophila
625	10.9	5	GPDA_DROEZ	1	349	Q27567	drosophila
626	10.9	5	ISPG_CLOPE	1	349	P58667	clostridium
627	10.9	5	AROQ_ECOLI	1	350	P00886	escherichia
628	10.9	5	OSTG_YEAST	1	350	P48439	saccharomyc
629	10.9	5	SUB1_SYNP7	1	350	P27366	synechococc
630	10.9	5	YCJQ_ECOLI	1	350	P76043	escherichia
631	10.9	5	YCXA_EUGGR	1	350	P31561	euglena gra
632	10.9	5	VMAT_P13B	1	351	P06166	bovine para
633	10.9	5	GPDA_DROVI	1	352	P07735	drosophila
634	10.9	5	KITH_HSVMD	1	352	P17653	marek's dis
635	10.9	5	OCT3_MOUSE	1	352	P20263	mus musculu
636	10.9	5	SUB1_SYNY3	1	352	Q01903	synechocyst
637	10.9	5	ACT_ACECL	1	353	P53491	acetabulari
638	10.9	5	GPDA_DROPS	1	353	Q27928	drosophila
639	10.9	5	HIS7_BUCAI	1	353	P57203	buchnera ap
640	10.9	5	HMDY_METJA	1	353	Q58734	methanococc
641	10.9	5	OP28_HAEIN	1	353	Q48221	haemophilus
642	10.9	5	SB11_CAEEL	1	353	P46506	caenorhabdi
643	10.9	5	T2BA_BACAR	1	353	P19887	bacillus an
644	10.9	5	VMAT_P13H4	1	353	P07873	human parai
645	10.9	5	ALF_CAMJE	1	354	P53818	campylobact
646	10.9	5	GLN1_ARATH	1	354	Q91v18	arabidopsis
647	10.9	5	PON1_MOUSE	1	354	P52430	mus musculu
648	10.9	5	GLN1_PEA	1	355	P08282	pisum sativ
649	10.9	5	GLN1_SOYBN	1	355	P24099	glycine max
650	10.9	5	GLN4_MAIZE	1	355	P38562	zea mays (m
651	10.9	5	GLN3_ORYSA	1	355	P14656	oryza sativ
652	10.9	5	NIFS_LACDE	1	355	P31672	lactobacill
653	10.9	5	Y816_DROME	1	356	Q9vaf0	drosophila
654	10.9	5	DXR_CAMJE	1	356	Q9pmv3	campylobact
655	10.9	5	GLN1_ALNGL	1	356	O04867	alnus gluti
656	10.9	5	GLN1_LOTJA	1	356	Q42899	lotus japon
657	10.9	5	GLN1_MEDSA	1	356	P04078	medicago sa
658	10.9	5	GLN1_PHAVU	1	356	P04770	phaseolus v
659	10.9	5	GLN2_VITVI	1	356	P51119	vitis vinif
660	10.9	5	GLN3_HORVU	1	356	Q06378	hordeum vul
661	10.9	5	GLN4_MAIZE	1	356	P38561	zea mays (m
662	10.9	5	GLN3_ORYSA	1	356	P14656	oryza sativ
663	10.9	5	GLNA_NICPL	1	356	P12424	nicotiana p
664	10.9	5	GLNA_VIGAC	1	356	P32289	vigna aconi
665	10.9	5	ACT1_OXYFA	1	357	P53503	oxytricha f
666	10.9	5	ACTC_SOLTU	1	357	P02583	oxytricha f
667	10.9	5	GLNA_PINSY	1	357	P30172	solanum tub
668	10.9	5	LCP5_YEAST	1	357	P52783	pinus sylve
669	10.9	5	METX_HAEIN	1	357	P40079	saccharomyc
670	10.9	5	YCIL_HAEIN	1	357	P45131	haemophilus
671	10.9	5	YMR2_EBV	1	357	P45104	haemophilus
672	10.9	5	ALF_BUCAI	1	358	P03192	epstein-bar
673	10.9	5	ALF_ECOLI	1	358	P57526	buchnera ap
674	10.9	5	ALF_EDWIC	1	358	P11604	escherichia
675	10.9	5	ALF_SCHPO	1	358	O52402	edwardsiell
676	10.9	5	ALF_YEAST	1	358	P36580	schizosacch
677	10.9	5	GLNA_LACSA	1	358	P14540	saccharomyc
678	10.9	5	ALF_BORBU	1	359	P23712	lactuca sat
679	10.9	5	ALF_HAEIN	1	359	O51401	borrelia bu
680	10.9	5	GPDA_DROKA	1	359	P44429	haemophilus
681	10.9	5	OP21_HAEIN	1	359	O97463	drosophila
682	10.9	5	RS2_SPICI	1	359	P43839	haemophilus
683	10.9	5	ALF_DROME	1	360	P19679	spiroplasma
684	10.9	5	DCAM_SOLTU	1	360	P07764	drosophila
685	10.9	5	OP26_HAEIN	1	360	Q04694	solanum tub
686	10.9	5	DCAM_NICSY	1	361	Q48216	haemophilus
687	10.9	5	DCAM_TOBAC	1	361	O80402	nicotiana s
688	10.9	5	ELYA_BACHD	1	361	O04009	nicotiana t
689	10.9	5	OP22_HAEIN	1	361	P41363	bacillus ha
690	10.9	5	OP29_HAEIN	1	361	Q20149	haemophilus
						Q48217	haemophilus

#			pg		ID	Organism	Accession
691	5	10.9	362	1	DCAM_DATST	datura stra	Q96555
692	5	10.9	362	1	NSDL_MOUSE	mus musculu	Q9rlj0
693	5	10.9	363	1	GPDA_DROME	drosophila	P17706
694	5	10.9	363	1	OP2B_HAEIN	haemophilus	Q48220
695	5	10.9	363	1	ORC2_ARATH	arabidopsis	Q38899
696	5	10.9	364	1	ATPG_SPIOL	spinacia ol	P05435
697	5	10.9	364	1	DNAJ_BORBU	borrelia bu	P28616
698	5	10.9	364	1	FCT1_HUMAN	homo sapien	Q96a29
699	5	10.9	364	1	YBR5_YEAST	saccharomyc	P38241
700	5	10.9	365	1	RF2_ECOLI	escherichia	P07012
701	5	10.9	365	1	RF2_HAEIN	haemophilus	P43918
702	5	10.9	365	1	RF2_SALTY	salmonella	P28353
703	5	10.9	366	1	ACTD_PHYPO	physarum po	P24263
704	5	10.9	366	1	ALF_NEUCR	neurospora	P53444
705	5	10.9	366	1	MRAY_RHIME	rhizobium m	Q52952
706	5	10.9	366	1	YNIO_YEAST	saccharomyc	P53938
707	5	10.9	367	1	OP2A_HAEIN	haemophilus	Q48219
708	5	10.9	367	1	TISD_MOUSE	mus musculu	P23949
709	5	10.9	368	1	ISPG_BUCAI	buchnera ap	P57374
710	5	10.9	369	1	METB_HAEIN	haemophilus	P44502
711	5	10.9	369	1	OP27_HAEIN	haemophilus	Q48218
712	5	10.9	369	1	PROB_NEIMA	neisseria m	Q9juk7
713	5	10.9	369	1	PROB_NEIMB	neisseria m	Q9izg2
714	5	10.9	369	1	RF2_THEMA	thermotoga	Q9xlr5
715	5	10.9	370	1	TRMU_BACSU	bacillus su	O35020
716	5	10.9	370	1	YXER_BACSU	bacillus su	P54957
717	5	10.9	371	1	ACT2_NAEFO	naegleria f	P27132
718	5	10.9	371	1	ACT5_DIPDE	diphyllobot	P53458
719	5	10.9	371	1	OP25_HAEIN	haemophilus	P46027
720	5	10.9	371	1	RQA1_HUMAN	homo sapien	P09651
721	5	10.9	371	1	TRMU_BACHD	bacillus ha	Q9kdf2
722	5	10.9	371	1	YI40_SYNY3	synechocyst	P73408
723	5	10.9	372	1	ACT_LUMRU	lumbricus r	P91754
724	5	10.9	373	1	ACT_CHOCR	chondrus cr	P53499
725	5	10.9	373	1	BIOF_HELPJ	helicobacte	Q9zln3
726	5	10.9	373	1	BIOF_HELPY	helicobacte	O25320
727	5	10.9	373	1	DMT1_HUMAN	homo sapien	Q9y5r6
728	5	10.9	373	1	DP3B_MYCPU	mycoplasma	Q98rk6
729	5	10.9	373	1	NSDL_HUMAN	homo sapien	Q15738
730	5	10.9	373	1	RO31_XENLA	xenopus lae	P51968
731	5	10.9	373	1	US74_DROME	drosophila	Q95y15
732	5	10.9	374	1	ACTM_STRPU	strongyloce	P12431
733	5	10.9	374	1	YF18_METJA	methanococc	Q58913
734	5	10.9	375	1	ACT1_ACACA	acanthamoeb	P02578
735	5	10.9	375	1	ACT1_DICDI	dictyosteli	P02577
736	5	10.9	375	1	ACT1_ECHGR	echinococcu	P35432
737	5	10.9	375	1	ACT1_FUGRU	fugu rubrip	P53484
738	5	10.9	375	1	ACT1_MAIZE	zea mays (m	P02582
739	5	10.9	375	1	ACT1_NAEFO	naegleria f	P27131
740	5	10.9	375	1	ACT1_OXYNO	oxytricha n	P12715
741	5	10.9	375	1	ACT1_OXYTR	oxytricha t	P53468
742	5	10.9	375	1	ACT1_PNECA	pneumocysti	P43239
743	5	10.9	375	1	ACT1_SCHCO	schizophyll	Q9y702
744	5	10.9	375	1	ACT1_SUIBO	suillus bov	Q9y701
745	5	10.9	375	1	ACT1_TETTH	tetrahymena	P10992
746	5	10.9	375	1	ACT2_DICDI	dictyosteli	P07827
747	5	10.9	375	1	ACT2_FUGRU	fugu rubrip	P53485
748	5	10.9	375	1	ACT2_OXYNO	oxytricha n	P55805
749	5	10.9	375	1	ACT2_OXYTR	oxytricha t	P53469
750	5	10.9	375	1	ACT2_SCHCO	schizophyll	Q9y896
751	5	10.9	375	1	ACT3_DICDI	dictyosteli	Q9y707
752	5	10.9	375	1	ACT3_FUGRU	fugu rubrip	P07829
753	5	10.9	375	1	ACT8_DICDI	dictyosteli	P07830
754	5	10.9	375	1	ACTA_PHYPO	physarum po	P02576
755	5	10.9	375	1	ACTB_CRIGR	cricetulus	P48975
756	5	10.9	375	1	ACTB_CYPCA	cyprinus ca	P12714
757	5	10.9	375	1	ACTB_HUMAN	homo sapien	P02570
758	5	10.9	375	1	ACTB_RABIT	oryctolagus	P29751
759	5	10.9	375	1	ACTB_SALSA	salmo salar	O42161
760	5	10.9	375	1	ACTC_BRABE	branchiosto	Q93129
761	5	10.9	375	1	ACTC_BRAFL	branchiosto	Q93131
762	5	10.9	375	1	ACTC_BRALA	branchiosto	O17503
764	5	10.9	375	1	ACTC_STYPL	styela plic	Q00215
765	5	10.9	375	1	ACTG_CEPAC	cephalospor	Q9uvw9
766	5	10.9	375	1	ACTG_EMENI	emericella	P20359
767	5	10.9	375	1	ACTG_HUMAN	homo sapien	P02571
768	5	10.9	375	1	ACTG_PENCH	penicillium	Q9urs0
769	5	10.9	375	1	ACT_AJECA	ajellomyces	P53455
770	5	10.9	375	1	ACT_BOTCI	botrytis ci	O13419
771	5	10.9	375	1	ACT_COPCI	coprinus ci	Q9uvx4
772	5	10.9	375	1	ACT_CRYNE	cryptococcu	P48465
773	5	10.9	375	1	ACT_FUCDI	fucus disti	P53502
774	5	10.9	375	1	ACT_NEUCR	neurospora	P78711
775	5	10.9	375	1	ACT_PHARH	phaffia rho	P53689
776	5	10.9	375	1	ACT_PUCGR	puccinia gr	P50138
777	5	10.9	375	1	ACT_SCHPO	schizosacch	P10989
778	5	10.9	375	1	ACT_THELA	thermomyces	P10365
779	5	10.9	375	1	DEK_HUMAN	homo sapien	P35659
780	5	10.9	375	1	DNAJ_ECOLI	escherichia	P08622
781	5	10.9	375	1	PER_DROCP	drosophila	P91607
782	5	10.9	375	1	PER_DROSC	drosophila	P91705
783	5	10.9	376	1	ACT1_AEDAE	aedes aegyp	P49128
784	5	10.9	376	1	ACT1_ARTSX	artemia sp.	P18600
785	5	10.9	376	1	ACT1_BOMMO	bombyx mori	P07836
786	5	10.9	376	1	ACT1_CAEEL	caenorhabdi	P10983
787	5	10.9	376	1	ACT1_DROME	drosophila	P10987
788	5	10.9	376	1	ACT1_HELER	heliocidari	P53462
789	5	10.9	376	1	ACT1_LUMTE	lumbricus t	P92182
790	5	10.9	376	1	ACT1_LYTPI	lytechinus	P53465
791	5	10.9	376	1	ACT1_ONCVO	onchocerca	P30162
792	5	10.9	376	1	ACT1_PEA	pisum sativ	P30164
793	5	10.9	376	1	ACT1_PHYIN	phytophthor	P22131
794	5	10.9	376	1	ACT1_PLAFA	plasmodium	P10988
795	5	10.9	376	1	ACT1_PODCA	podocoryne	P41112
796	5	10.9	376	1	ACT1_SACKO	saccoglossu	O18499
797	5	10.9	376	1	ACT1_SCHMA	schistosoma	P53470
798	5	10.9	376	1	ACT1_STRFN	strongyloce	P10990
799	5	10.9	376	1	ACT2_ARTSX	artemia sp.	P18601
800	5	10.9	376	1	ACT2_BACDO	bactrocera	P45885
801	5	10.9	376	1	ACT2_BOMMO	bombyx mori	P07837
802	5	10.9	376	1	ACT2_CAEEL	caenorhabdi	P10984
803	5	10.9	376	1	ACT2_DIPDE	diphyllobot	P53456
804	5	10.9	376	1	ACT2_DROME	drosophila	P02572
805	5	10.9	376	1	ACT2_ECHGR	echinococcu	Q03341
806	5	10.9	376	1	ACT2_LUMTE	lumbricus t	P92176
807	5	10.9	376	1	ACT2_LYTPI	lytechinus	P53466
808	5	10.9	376	1	ACT2_ONCVO	onchocerca	P30163
809	5	10.9	376	1	ACT2_PEA	pisum sativ	P30165
810	5	10.9	376	1	ACT2_PLAFA	plasmodium	P14883
811	5	10.9	376	1	ACT2_SACKO	saccoglossu	O18500
812	5	10.9	376	1	ACT2_SCHMA	schistosoma	P53471
813	5	10.9	376	1	ACT2_TETPY	tetrahymena	P10993
814	5	10.9	376	1	ACT3_BACDO	bactrocera	P45886
815	5	10.9	376	1	ACT3_BOMMO	bombyx mori	P04829
816	5	10.9	376	1	ACT3_DROME	drosophila	P53501
817	5	10.9	376	1	ACT3_HELAM	helicoverpa	Q25010
818	5	10.9	376	1	ACT3_LIMPO	limulus pol	P41340
819	5	10.9	376	1	ACT3_ORYSA	oryza sativ	P17299
820	5	10.9	376	1	ACT3_PODCA	podocoryne	P41113
821	5	10.9	376	1	ACT3_SOYBN	glycine max	P02580
822	5	10.9	376	1	ACT4_ARTSX	artemia sp.	P18603
823	5	10.9	376	1	ACT4_BOMMO	bombyx mori	Q27250
824	5	10.9	376	1	ACT4_CAEEL	caenorhabdi	P10986
825	5	10.9	376	1	ACT4_DROME	drosophila	P02574
826	5	10.9	376	1	ACT5_BACDO	bactrocera	P45887
827	5	10.9	376	1	ACT5_CHICK	gallus gall	P53478
828	5	10.9	376	1	ACT5_DROME	drosophila	P10981
829	5	10.9	376	1	ACT5_XENLA	xenopus lae	P53505
830	5	10.9	376	1	ACT6_DROME	drosophila	P02575
831	5	10.9	376	1	ACT7_ORYSA	oryza sativ	P17300
832	5	10.9	376	1	ACT8_XENLA	xenopus lae	P53506
833	5	10.9	376	1	ACTA_LIMPO	limulus pol	P41339
834	5	10.9	376	1	ACTB_STRPU	strongyloce	P53472
835	5	10.9	376	1	ACTB_STRPU	strongyloce	P53473
836	5	10.9	376	1	ACTC_HALRO	halocynthia	P53461

#	len	sc	w	acc	species	name	n
837	376	10.9	5	P12716	pisaster oc	ACTC_PISOC	1
838	376	10.9	5	Q07903	strongyloce	ACTC_STRPU	1
839	376	10.9	5	O10991	molgula ocu	ACTE_STRPU	1
840	376	10.9	5	P53474	strongyloce	ACTE_STRPU	1
841	376	10.9	5	P18499	strongyloce	ACTF_STRPU	1
842	376	10.9	5	P12718	homo sapien	ACTH_HUMAN	1
843	376	10.9	5	P17304	aplysia cal	ACTM_APLCA	1
844	376	10.9	5	P53463	heliocidari	ACTM_HELER	1
845	376	10.9	5	P53464	heliocidari	ACTM_HELTB	1
846	376	10.9	5	P12717	pisaster oc	ACTM_PISOC	1
847	376	10.9	5	P53483	fugu rubrip	ACTX_FUGRU	1
848	376	10.9	5	P41341	limulus pol	ACTY_LIMPO	1
849	376	10.9	5	P26182	achlya bise	ACT_ACHBI	1
850	376	10.9	5	P92179	biomphalari	ACT_BIOGL	1
851	376	10.9	5	P90689	brugia mala	ACT_BRUMA	1
852	376	10.9	5	P14235	candida alb	ACT_CANAL	1
853	376	10.9	5	Q9uz8	candida dub	ACT_CANDU	1
854	376	10.9	5	O17320	crassostrea	ACT_CRAGI	1
855	376	10.9	5	P26183	cryptospori	ACT_CRYPV	1
856	376	10.9	5	P11426	entamoeba h	ACT_ENTHI	1
857	376	10.9	5	Q39758	fucus vesic	ACT_FUCVE	1
858	376	10.9	5	P17126	hydra atten	ACT_HYDAT	1
859	376	10.9	5	P49871	manduca sex	ACT_MANSE	1
860	376	10.9	5	O16808	mayetiola d	ACT_MAYDE	1
861	376	10.9	5	O74258	pichia angu	ACT_PICAN	1
862	376	10.9	5	Q26065	placopecten	ACT_PLAMG	1
863	376	10.9	5	P14227	taenia soli	ACT_TAESO	1
864	376	10.9	5	P53476	toxoplasma	ACT_TOXGO	1
865	376	10.9	5	O43502	homo sapien	R51C_HUMAN	1
866	377	10.9	5	P42442	corynebacte	RECA_CORGL	1
867	377	10.9	5	P49942	bacteroides	XYNA_BACOV	1
868	377	10.9	5	Q98972	oryzias lat	ACT1_ORYLA	1
869	377	10.9	5	P13362	oryza sativ	ACT1_ORYSA	1
870	377	10.9	5	P53504	sorghum bic	ACT1_SORBI	1
871	377	10.9	5	P02581	glycine max	ACT1_SOYBN	1
872	377	10.9	5	Q05214	nicotiana t	ACT1_TOBAC	1
873	377	10.9	5	P04751	xenopus lae	ACT1_XENLA	1
874	377	10.9	5	P26197	absidia gla	ACT2_ABSGL	1
875	377	10.9	5	P53492	arabidopsis	ACT2_ARATH	1
876	377	10.9	5	P10995	xenopus lae	ACT2_XENLA	1
877	377	10.9	5	P20399	xenopus tro	ACT2_XENTR	1
878	377	10.9	5	P53493	arabidopsis	ACT3_ARATH	1
879	377	10.9	5	P53457	diphyllobot	ACT3_DIPDE	1
880	377	10.9	5	P46258	pisum sativ	ACT3_PEA	1
881	377	10.9	5	P30167	solanum tub	ACT3_SOLTU	1
882	377	10.9	5	P04752	xenopus lae	ACT3_XENLA	1
883	377	10.9	5	P53494	arabidopsis	ACT4_ARATH	1
884	377	10.9	5	P30168	solanum tub	ACT6_SOLTU	1
885	377	10.9	5	P30169	solanum tub	ACT7_SOLTU	1
886	377	10.9	5	P08023	gallus gall	ACTA_CHICK	1
887	377	10.9	5	P53479	homo sapien	ACTA_HUMAN	1
888	377	10.9	5	P53496	arabidopsis	ACTB_ARATH	1
889	377	10.9	5	P30171	solanum tub	ACTB_SOLTU	1
890	377	10.9	5	P53497	arabidopsis	ACTC_ARATH	1
891	377	10.9	5	P53480	fugu rubrip	ACTC_FUGRU	1
892	377	10.9	5	P04270	homo sapien	ACTC_HUMAN	1
893	377	10.9	5	P30173	solanum tub	ACTD_SOLTU	1
894	377	10.9	5	P49055	carassius a	ACTS_CARAU	1
895	377	10.9	5	P53481	cyprinus ca	ACTS_CYPCA	1
896	377	10.9	5	P02568	fugu rubrip	ACTS_FUGRU	1
897	377	10.9	5	P53482	homo sapien	ACTS_HUMAN	1
898	377	10.9	5	P53498	fugu rubrip	ACTT_FUGRU	1
899	377	10.9	5	O65315	chlamydomon	ACT_CHLRE	1
900	377	10.9	5	P53500	coleochaete	ACT_COLSC	1
901	377	10.9	5	O65316	cyanidiosch	ACT_CYAME	1
902	377	10.9	5	P20904	mesostigma	ACT_MESVI	1
903	377	10.9	5	O32465	volvox cart	ACT_VOLCA	1
904	377	10.9	5	P90680	buchnera ap	DNAJ_BUCAI	1
905	377	10.9	5	P03955	apis mellif	OPSB_APIME	1
906	377	10.9	5	P53460	macaca fusc	PEPC_MACFU	1
907	377	10.9	5	P27130	halocynthia	ACT1_HALRO	1
908	378	10.9	5	Q25472	halocynthia	ACT2_HALRO	1
909	378	10.9	5		molgula ocu	ACT2_MOLOC	1
910	378	10.9	5	Q93132	branchiosto	ACTM_BRAFL	1
911	378	10.9	5	O15998	ciona savig	ACTM_CIOSA	1
912	378	10.9	5	P53467	molgula ocu	ACTM_MOLOC	1
913	378	10.9	5	P26198	styela clav	ACTM_STYCL	1
914	378	10.9	5	P53475	styela clav	ACTN_STYCL	1
915	378	10.9	5	O65314	scherffelia	ACT1_SCHDU	1
916	378	10.9	5	Q60004	salmonella	DNAJ_SALTY	1
917	378	10.9	5	O06672	streptococc	DP3B_STRPN	1
918	378	10.9	5	P81914	drosophila	O33A_DROME	1
919	378	10.9	5	P28917	escherichia	YDCC_ECOLI	1
920	378	10.9	5	P28912	escherichia	YHHI_ECOLI	1
921	378	10.9	5	P46080	anabaena sp	YN91_ANASP	1
922	379	10.9	5	P17298	oryza sativ	ACT2_ORYSA	1
923	379	10.9	5	Q93110	branchiosto	ACTM_BRABE	1
924	379	10.9	5	Q00214	styela plic	ACTM_STYPL	1
925	379	10.9	5	O50161	borrelia bu	Y374_BORBU	1
926	380	10.9	5	P23343	daucus caro	ACT1_DAUCA	1
927	380	10.9	5	O17502	branchiosto	ACTM_BRALA	1
928	380	10.9	5	Q45552	bacillus st	DNAJ_BACST	1
929	380	10.9	5	Q9zmw7	helicobacte	METB_HELPJ	1
930	380	10.9	5	P56069	helicobacte	METB_HELPY	1
931	380	10.9	5	Q00374	colletotric	PLYA_COLGL	1
932	380	10.9	5	P32652	maize dwarf	POLG_MDMV	1
933	381	10.9	5	P23344	daucus caro	ACT2_DAUCA	1
934	382	10.9	5	P03090	mouse polyo	COA1_POVMA	1
935	382	10.9	5	P38494	bacillus su	RSIH_BACSU	1
936	383	10.9	5	P03091	mouse polyo	COA1_POVM3	1
937	383	10.9	5	P12907	mouse polyo	COA1_POVMC	1
938	383	10.9	5	P49302	mouse polyo	COA1_POVMP	1
939	385	10.9	5	P42023	pneumocysti	ACT2_PNECA	1
940	385	10.9	5	Q04064	bordetella	OMP_BORPE	1
941	385	10.9	5	P46025	haemophilus	OP23_HAEIN	1
942	385	10.9	5	P91686	drosophila	PER_DRONE	1
943	385	10.9	5	P47702	mycoplasma	Y464_MYCGE	1
944	386	10.9	5	P46026	haemophilus	OP24_HAEIN	1
945	387	10.9	5	Q9eqn3	mus musculu	TIZ2_MOUSE	1
946	388	10.9	5	Q9rgs1	staphylococ	GAL1_STACA	1
947	388	10.9	5	Q9n2d3	callithrix	PEPC_CALJA	1
948	388	10.9	5	P20142	homo sapien	PEPC_HUMAN	1
949	388	10.9	5	P30758	streptococc	RECA_STRPN	1
950	389	10.9	5	P22806	bacillus sp	BIOF_BACSH	1
951	389	10.9	5	P35515	methanosarc	DNAJ_METMA	1
952	390	10.9	5	O51752	borrelia bu	DFP_BORBU	1
953	390	10.9	5	P23256	escherichia	MALY_ECOLI	1
954	390	10.9	5	P91716	drosophila	PER_DROTP	1
955	390	10.9	5	P19717	mumps virus	RRPP_MUMP1	1
956	390	10.9	5	P55642	rhizobium s	Y4RI_RHISN	1
957	391	10.9	5	P25796	cavia porce	CATE_CAVPO	1
958	391	10.9	5	Q981f7	rhizobium l	EFTU_RHILO	1
959	391	10.9	5	Q9r0n0	mus musculu	GAL1_MOUSE	1
960	391	10.9	5	P91613	drosophila	PER_DROIN	1
961	391	10.9	5	P16072	mumps virus	RRPP_MUMPE	1
962	392	10.9	5	P16595	mumps virus	RRPP_MUMPM	1
963	392	10.9	5	P79106	bos taurus	PAF2_BOVIN	1
964	392	10.9	5	Q99487	homo sapien	PAF2_HUMAN	1
965	392	10.9	5	Q96p15	buchnera ap	SB11_HUMAN	1
966	392	10.9	5	Q59169	buchnera ap	TRPB_BUCSC	1
967	393	10.9	5	O83475	treponema p	Y462_TREPA	1
968	393	10.9	5	P47599	mycoplasma	ACKA_MYCGE	1
969	393	10.9	5	P01869	mus musculu	GC1M_MOUSE	1
970	393	10.9	5	O25443	helicobacte	MT04_HELPY	1
971	393	10.9	5	Q59181	borrelia bu	PGK_BORBU	1
972	394	10.9	5	O14649	homo sapien	CIW3_HUMAN	1
973	394	10.9	5	Q13115	homo sapien	DUS4_HUMAN	1
974	394	10.9	5	P39642	bacillus su	YWFF_BACSU	1
975	395	10.9	5	O53598	streptococc	NEUA_STRAG	1
976	395	10.9	5	P41999	caenorhabdi	NH10_CAEEL	1
977	395	10.9	5	P92203	drosophila	PER_DROEQ	1
978	395	10.9	5	Q9y3q8	homo sapien	TIZ2_HUMAN	1
979	395	10.9	5	O57857	methanococc	Y414_METJA	1
980	396	10.9	5	P06727	homo sapien	APA4_HUMAN	1
981	396	10.9	5	P25023	ratus norv	BRB2_RAT	1
982	396	10.9	5	P43159	oryctolagus	CATE_RABIT	1

```
983    5  10.9  PER_DROPU   396  1    P91697  drosophila
984    5  10.9  PER_DROPV   396  1    P91698  drosophila
985    5  10.9  PGK_STAAM   396  1    Q925c4  staphylococc
986    5  10.9  SOTB_HAEIN  396  1    P44535  haemophilus
987    5  10.9  YD18_YEAST  396  1    Q12185  saccharomyc
988    5  10.9  YWBD_BACSU  396  1    P39587  bacillus su
989    5  10.9  6P22_YEAST  397  1    Q12471  saccharomyc
990    5  10.9  CATE_MOUSE  397  1    P70269  mus musculu
991    5  10.9  CD4_ERYPA   397  1    Q08339  erythrocebu
992    5  10.9  DJA4_MOUSE  397  1    Q9jmc3  mus musculu
993    5  10.9  LHX3_HUMAN  397  1    Q9ubr4  homo sapien
994    5  10.9  THIM_HUMAN  397  1    P42765  homo sapien
995    5  10.9  THIM_RAT    397  1    P13437  rattus norv
996    5  10.9  APL1_HUMAN  398  1    O14791  homo sapien
997    5  10.9  CATE_RAT    398  1    P16228  rattus norv
998    5  10.9  DAP3_HUMAN  398  1    P51398  homo sapien
999    5  10.9  TRA5_RHIME  398  1    Q52873  rhizobium m
1000   5  10.9  GCAM_MOUSE  399  1    P01865  mus musculu
```

ALIGNMENTS

RESULT 1
Y228_BORBU
ID   Y228_BORBU   STANDARD;   PRT;   971 AA.
AC   O51246;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein BB0228.
GN   BB0228.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria: Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).

DR   EMBL; AE001133; AAC66621.1; -.
DR   TIGR; BB0228; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   971 AA; 112959 MW; 088A688D7BBC591A CRC64;

Query Match   19.6%; Score 9; DB 1; Length 971;
Best Local Similarity   100.0%; Pred. No. 0.16;
Matches   9; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   4 TLSLKNDFK 12
           |||||||||
Db   653 TLSLKNDFK 661

RESULT 2
FIL1_YEAST
ID   FIL1_YEAST   STANDARD;   PRT;   230 AA.
AC   P38771;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   FIL1 protein, mitochondrial precursor.
GN   FIL1 OR KIM4 OR YHR038W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RC   STRAIN=MT8-1;
RX   MEDLINE=98417448; PubMed=9746366;
RA   Kanai T., Takeshita S., Atomi H., Umemura K., Ueda M., Tanaka A.;
RT   "A regulatory factor, Fil1p, involved in derepression of the
RT   isocitrate lyase gene in Saccharomyces cerevisiae -- a possible
RT   mitochondrial protein necessary for protein synthesis in
RT   mitochondria.";
RL   Eur. J. Biochem. 256:212-220(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [3]
RP   SIMILARITY TO RRF.
RX   MEDLINE=96155066; PubMed=8563640;
RA   Ouzounis C., Bork P., Casari G., Sander C.;
RT   "New protein functions in yeast chromosome VIII.";
RL   Protein Sci. 4:2424-2428(1995).
CC   -!- FUNCTION: INVOLVED IN GLUCOSE REPRESSION/DEREPRESSION. MAY BE
CC   NECESSARY FOR PROTEIN SYNTHESIS IN MITOCHONDRIA. MAY FUNCTION AS A
CC   RIBOSOME RECYCLING FACTOR IN MITOCHONDRIA.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE RRF FAMILY.

DR   EMBL; AB016033; BAA31687.1; -.
DR   EMBL; U00062; AAB68906.1; -.
DR   PIR; S46737; S46737.
DR   SGD; S0001080; FIL1.
DR   InterPro; IPR002661; RRF.
DR   Pfam; PF01765; RRF; 1.
KW   Protein biosynthesis; Transit peptide; Mitochondrion.
FT   TRANSIT   1   24   MITOCHONDRION (POTENTIAL).
FT   CHAIN   25   230   FIL1 PROTEIN.
SQ   SEQUENCE   230 AA; 26406 MW; 9CBCD8C5F86F3008 CRC64;

Query Match   15.2%; Score 7; DB 1; Length 230;
Best Local Similarity   100.0%; Pred. No. 5.1;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   34 RLFNRSF 40

```
Db      17 RLFNRSF 23
           |||||||
```

```
RESULT 3
ABF1_TRIRE
ID   ABF1_TRIRE     STANDARD;      PRT;   500 AA.
AC   Q92455;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN   ABF1.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=QM9414 / RUT C-30;
RX   MEDLINE=96434473; PubMed=8837440;
RA   Margolles-Clark E., Tenkanen M., Nakari-Setaelae T., Penttilae M.;
RT   "Cloning of genes encoding alpha-L-arabinofuranosidase and beta-
RT   xylosidase from Trichoderma reesei by expression in Saccharomyces
RT   cerevisiae.";
RL   Appl. Environ. Microbiol. 62:3840-3846(1996).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC       arabinofuranoside residues in alpha-L-arabinosides.
CC   -!- PATHWAY: INVOLVED IN DEGRADATION OF THE PLANT CELL WALL
CC       POLYSACCHARIDE L-ARABINAN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; 269252; CAA93243.1; -.
KW   Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    500       ALPHA-L-ARABINOFURANOSIDASE.
FT   CARBOHYD    467    467       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   500 AA;  51115 MW;  FC6B4DB03EE3C762 CRC64;

Query Match             15.2%;  Score 7;  DB 1;  Length 500;
Best Local Similarity  100.0%;  Pred. No. 10;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 SGSLSTF 32
           |||||||
Db     266 SGSLSTF 272


RESULT 4
XYL1_TRIKO
ID   XYL1_TRIKO     STANDARD;      PRT;   500 AA.
AC   P48792;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Arabinofuranosidase/B-xylosidase precursor [Includes: Alpha-L-
DE   arabinofuranosidase (EC 3.2.1.55) (Arabinosidase); Beta-xylosidase
DE   (EC 3.2.1.37) (1,4-beta-D-xylan xylohydrolase) (Xylan 1,4-beta-
DE   xylosidase)].
GN   XYL1.
OS   Trichoderma koningii.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=55202;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=G-39;
RA   Huang L.N., Hseu T.H., Lee Y.J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC       arabinofuranoside residues in alpha-L-arabinosides.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-XYLANS SO AS TO
CC       REMOVE SUCCESSIVE D-XYLOSE RESIDUES FROM THE NON-REDUCING
CC       TERMINI. IT ALSO HYDROLYSES XYLOBIOSE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U38661; AAA81024.1; -.
KW   Xylan degradation; Hydrolase; Glycosidase; Multifunctional enzyme;
KW   Glycoprotein; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    500       ARABINOFURANOSIDASE/B-XYLOSIDASE.
FT   CARBOHYD    467    467       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   500 AA;  51129 MW;  06DFC319AAFA1149 CRC64;

Query Match             15.2%;  Score 7;  DB 1;  Length 500;
Best Local Similarity  100.0%;  Pred. No. 10;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 SGSLSTF 32
           |||||||
Db     266 SGSLSTF 272


RESULT 5
RELB_MOUSE
ID   RELB_MOUSE     STANDARD;      PRT;   558 AA.
AC   Q04863;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor RELB.
GN   RELB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92123192; PubMed=1732739;
RA   Ryseck R.P., Bull P., Takamiya M., Bours V., Siebenlist U.,
RA   Dobrzanski P., Bravo R.;
RT   "RelB, a new Rel family transcription activator that can interact
RT   with p50-NF-kappa B.";
RL   Mol. Cell. Biol. 12:674-684(1992).
RN   [2]
RP   SEQUENCE OF 1-116 FROM N.A.
RX   MEDLINE=93180804; PubMed=8441398;
RA   Dobrzanski P., Ryseck R.P., Bravo R.;
RT   "Both N- and C-terminal domains of RelB are required for full
RT   transactivation: role of the N-terminal leucine zipper-like motif.";
RL   Mol. Cell. Biol. 13:1572-1582(1993).
RN   [3]
RP   SEQUENCE OF 309-429 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RX   MEDLINE=95147976; PubMed=7845467;
RA   Burkly L., Hession C., Ogata I., Reilly C., Marconi L.A.,
RA   Olson D., Tizard R., Cate R., Lo D.;
RT   "Expression of relB is required for the development of thymic medulla
RT   and dendritic cells.";
RL   Nature 373:531-536(1995).
```

```
CC   -!- FUNCTION: STIMULATES PROMOTER ACTIVITY IN THE PRESENCE OF P49- AND
CC       P50-NF-KAPPA-B. NEITHER ASSOCIATES WITH DNA NOR WITH P65-NF-KAPPA-
CC       B.
CC   -!- SUBCELLULAR LOCATION: Nuclear
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN INTESTINE, THYMUS AND SPLEEN.
CC       UNDETECTABLE IN LIVER, BOME MARROW, KIDNEY AND TESTIS.
CC   -!- DOMAIN: BOTH N- AND C-TERMINAL DOMAINS ARE REQUIRED FOR
CC       TRANSCRIPTIONAL ACTIVATION.
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83380; AAA40041.1; -.
DR   EMBL; S56076; AAB25493.2; -.
DR   EMBL; S76754; AAB33259.1; -.
DR   PIR; A42023; A42023.
DR   HSSP; P25799; 1BFT.
DR   TRANSFAC; T01932; -.
DR   MGD; MGI:103289; Relb.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; REL.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   PRINTS; PR00057; NFKBTNSCPFCT.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; 1.
DR   PROSITE; PS50254; REL_2; 1.
KW   Nuclear protein; Transcription regulation; Activator; Phosphorylation.
FT   DOMAIN        22     50       LEUCINE-ZIPPER.
FT   DOMAIN       103    418       REL-LIKE (RHD).
FT   DOMAIN       387    391       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN       411    416       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   CONFLICT      51     51       V -> D (IN REF. 2).
SQ   SEQUENCE   558 AA;  60304 MW;  1ED2A354C6ED0D3D CRC64;

Query Match             15.2%;  Score 7;  DB 1;  Length 558;
Best Local Similarity  100.0%;  Pred. No. 11;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     16 TDELEII 22
          |||||||
Db     33 TDELEII 39

RESULT 6
RELB_HUMAN
ID   RELB_HUMAN     STANDARD;      PRT;    579 AA.
AC   Q01201;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor RELB (I-Rel).
GN   RELB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   TISSUE=T-cell;
RX   MEDLINE=92249768; PubMed=1577270;
RA   Ruben S.M., Klement J.F., Maher M., Coleman T.A., Chen C.H.,
RA   Rosen C.A.;
RT   "I-Rel: a novel rel-related protein that inhibits NF-kappa B
RT   transcriptional activity.";
RL   Genes Dev. 6:745-760(1992).
RN   [2]
```

```
RP   FUNCTION
RX   MEDLINE=93180804; PubMed=8441398;
RA   Dobrzanski P., Ryseck R.P., Bravo R.;
RT   "Both N- and C-terminal domains of RelB are required for full
RT   transactivation: role of the N-terminal leucine zipper-like motif.";
RL   Mol. Cell. Biol. 13:1572-1582(1993).
CC   -!- FUNCTION: STIMULATES PROMOTER ACTIVITY IN THE PRESENCE OF P49- AND
CC       P50-NF-KAPPA-B. NEITHER ASSOCIATES WITH DNA NOR WITH P65-NF-KAPPA-
CC       B.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- INDUCTION: BY MITOGENS.
CC   -!- DOMAIN: BOTH N- AND C-TERMINAL DOMAINS ARE REQUIRED FOR
CC       TRANSCRIPTIONAL ACTIVATION.
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO INHIBIT THE
CC       TRANSCRIPTIONAL ACTIVITY OF NUCLEAR FACTOR NF-KAPPA-B.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83221; AAA36127.1; -.
DR   PIR; A42617; A42617.
DR   HSSP; P25799; 1BFT.
DR   TRANSFAC; T01931; -.
DR   MIM; 604758; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; REL.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   PRINTS; PR00057; NFKBTNSCPFCT.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; 1.
DR   PROSITE; PS50254; REL_2; 1.
KW   Nuclear protein; Transcription regulation; Activator; Phosphorylation.
FT   DOMAIN        40     68       LEUCINE-2IPPER.
FT   DOMAIN       125    440       REL-LIKE (RHD).
FT   DOMAIN       433    438       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   579 AA;  61990 MW;  5BFF8F8A5184D52 CRC64;

Query Match             15.2%;  Score 7;  DB 1;  Length 579;
Best Local Similarity  100.0%;  Pred. No. 11;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     16 TDELEII 22
          |||||||
Db     51 TDELEII 57

RESULT 7
YN48_YEAST
ID   YN48_YEAST     STANDARD;      PRT;    591 AA.
AC   P42846;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1996 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 68.7 kDa protein in STB1-MCK1 intergenic region.
GN   YNL308C OR NO388.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1676;
RX   MEDLINE=96076632; PubMed=7502583;
RA   Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;
RT   "Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV
RT   identifies six known genes, a new member of the hexose transporter
```

```
RT   family and ten new open reading frames.";
RL   Yeast 11:1077-1085(1995).
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC22G7.05.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z46259; CAA86387.1; -.
DR   EMBL; Z71584; CAA96237.1; -.
DR   SGD; S0005252; YNL308C.
KW   Hypothetical protein.
FT   DOMAIN       53     65       POLY-GLU.
FT   DOMAIN      457    465       POLY-GLU.
SQ   SEQUENCE   591 AA;  68653 MW;  3B63757lEDB05EA6 CRC64;

Query Match
Best Local Similarity   15.2%;  Score 7;  DB 1;  Length 591;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7  LKNDFKE  13
            |||||||
Db     352  LKNDFKE  358

RESULT 8
ID   UGS3_SOLTU     STANDARD;      PRT;    788 AA.
AC   Q43847;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Glycogen [starch] synthase, chloroplast precursor (EC 2.4.1.11)
DE   (GBSSII) (Granule-bound starch synthase II) (Fragment).
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae: euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 66-75.
RC   STRAIN=CV. DESIREE; TISSUE=Tuber;
RX   MEDLINE=95400340; PubMed=7670507;
RA   Edwards A., Marshall J., Sidebottom C., Visser R.G.F., Smith A.M.,
RA   Martin C.;
RT   "Biochemical and molecular characterization of a novel starch
RT   synthase from potato tubers.";
RL   Plant J. 8:283-294(1995).
CC   -!- FUNCTION: ACCOUNTS FOR ONLY 10 TO 15% OF THE TOTAL SOLUBLE STARCH
CC       SYNTHASE ACTIVITY IN TUBERS.
CC   -!- CATALYTIC ACTIVITY: UDP-glucose + [(1,4)-alpha-D-glucosyl](N) =
CC       UDP + [(1,4)-alpha-D-glucosyl](N+1).
CC   -!- PATHWAY: STARCH BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST/AMYLOPLAST; SOLUBLE AND GRANULE-
CC       BOUND.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE
CC       FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X87988; CAA61241.1; -.
DR   InterPro; IPR001296; Glycos_transf_1.
DR   Pfam; PF00534; Glycos_transf_1; 1.
```

```
KW   Glycogen biosynthesis; Transferase; Glycosyltransferase;
KW   Transit peptide; Chloroplast; Amyloplast; Starch biosynthesis.
FT   NON_TER       1      1                                 CHLOROPLAST.
FT   TRANSIT      <1     65                                 GLYCOGEN [STARCH] SYNTHASE.
FT   CHAIN        66    788                                 UDP-GLUCOSE (BY SIMILARITY).
FT.  BINDING     310    310                                 S -> D.
FT   VARIANT      71     71
SQ   SEQUENCE   788 AA;  87890 MW;  8D8B90611E862B7B CRC64;

Query Match
Best Local Similarity   15.2%;  Score 7;  DB 1;  Length 788;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24  GGSGSLS  30
            |||||||
Db     128  GGSGSLS  134

RESULT 9
ID   K6PP_RABIT     STANDARD;      PRT;    791 AA.
AC   P47859;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   6-phosphofructokinase, type C (EC 2.7.1.11) (Phosphofructokinase
DE   1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme C) (PFK-C).
GN   PFKP.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=94164929; PubMed=8119919;
RA   Li Y., Valaitis A.P., Latshaw S.P., Kwiatkowska D., Tripathi R.L.,
RA   Campbell M.C., Kemp R.G.;
RT   "Structure and expression of the cDNA for the C isozyme of
RT   phosphofructo-1-kinase from rabbit brain.";
RL   J. Biol. Chem. 269:5781-5787(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC       fructose 1,6-bisphosphate.
CC   -!- COFACTOR: MAGNESIUM.
CC   -!- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC       FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC   -!- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
CC       SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U01154; AAA17707.1; -.
DR   HSSP; P00512; 3PFK.
DR   InterPro; IPR000023; PFK; 4.
DR   Pfam; PF00365; PFK; 4.
DR   PRINTS; PR00476; PHFRCTKINASE.
DR   ProDom; PD007707; Phosphofructokinase; 2.
DR   PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
KW   Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW   Phosphorylation; Magnesium; Multigene family.
SQ   SEQUENCE   791 AA;  86349 MW;  3C10A36F229FD8E8 CRC64;

Query Match
Best Local Similarity   15.2%;  Score 7;  DB 1;  Length 791;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   23 IGGSGSL 29
         ||||||
Db  125 IGGSGSL 131

RESULT 10
PSD8_DICDI
ID   PSD8_DICDI     STANDARD;      PRT;     88 AA.
AC   P02889; P09409;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vegetative cell protein X (M4 protein) (Fragment).
GN   M4.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81124286; PubMed=7465423;
RA   Kimmel A.R., Firtel R.A.;
RT   "Intervening sequences in a Dictyostelium gene that encodes a low
RT   abundance class mRNA.";
RL   Nucleic Acids Res. 8:5599-5610(1980).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RX   MEDLINE=86310885; PubMed=2427932;
RA   Kimmel A.R., Firtel R.A.;
RT   "Sequence organization and developmental expression of an
RT   interspersed, repetitive element and associated single-copy DNA
RT   sequences in Dictyostelium discoideum.";
RL   Mol. Cell. Biol. 5:2123-2130(1985).
CC   -!- FUNCTION: ACTS AS A REGULATORY SUBUNIT OF THE 26 PROTEASOME WHICH
CC       IS INVOLVED IN THE ATP-DEPENDENT DEGRADATION OF UBIQUITINATED
CC       PROTEINS (BY SIMILARITY).
CC   -!- DEVELOPMENTAL STAGE: POTENTIALLY SUBJECT TO DEVELOPMENTAL
CC       REGULATION.
CC   -!- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT S14 FAMILY.
CC   ---------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------
DR   EMBL; V00194; CAA23485.1; -.
DR   EMBL; M11429; AAA33223.1; -.
DR   PIR; A03384; QXDO.
DR   DictyDb; DD01002; -.
KW   Proteasome; Developmental protein.
FT   NON_TER      88     88
SQ   SEQUENCE   88 AA;   10130 MW;   370B5997E882B853 CRC64;

Qy    3 KTLSLK 8
         ||||||
Db   83 KTLSLK 88

Query Match     13.0%;  Score 6;  DB 1;  Pred. No. 23;
Best Local Similarity   100.0%;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 11
CYC6_PORPU
ID   CYC6_PORPU     STANDARD;      PRT;     110 AA.
AC   P51200;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytochrome C6 precursor (Soluble cytochrome F) (Cytochrome C553).
GN   PETJ.
```

```
OS   Porphyra purpurea.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX   NCBI_TaxID=2787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RA   Reith M.E., Munholland J.;
RT   "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT   genome.";
RL   Plant Mol. Biol. Rep. 13:333-335(1995).
CC   -!- FUNCTION: CYTOCHROME C6 IS A MONOHEME MONOMER. IT FUNCTIONS AS AN
CC       ELECTRON CARRIER BETWEEN MEMBRANE-BOUND CYTOCHROME F AND P700 IN
CC       THE PHOTOPHOSPHORYLATION CHAIN IN CHLOROPLASTS AND ALGAE. IT
CC       SUBSTITUTES FOR PLASTOCYANIN IN COPPER-DEFICIENT BLUE-GREEN ALGAE
CC       AND IN THE CHLOROPLASTS OF SOME EUKARYOTE ALGAE.
CC   -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC   ---------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------
DR   EMBL; U38804; AAC08086.1; -.
DR   HSSP; P56534; 1C6S.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR003088; Cyt_CI.
DR   InterPro; IPR002329; Cyt_CIC.
DR   Pfam; PF00034; cytochrome_c; 1.
DR   PRINTS; PR00605; CYTCHROMECIC.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Electron transport; Chloroplast; Photosynthesis; Heme; Thylakoid;
KW   Signal.
FT   SIGNAL       1     25               BY SIMILARITY.
FT   CHAIN       26    110               CYTOCHROME C6.
FT   BINDING     39     39               HEME (COVALENT).
FT   BINDING     42     42               HEME (COVALENT).
FT   METAL       43     43               IRON (HEME AXIAL LIGAND).
FT   METAL       83     83               IRON (HEME AXIAL LIGAND).
SQ   SEQUENCE   110 AA;   11745 MW;   E94847B79168995A CRC64;

Qy    1 MKKTLS 6
         ||||||
Db    1 MKKTLS 6

Query Match     13.0%;  Score 6;  DB 1;  Length 110;
Best Local Similarity   100.0%;  Pred. No. 28;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 12
HV3B_HUMAN
ID   HV3B_HUMAN     STANDARD;      PRT;     114 AA.
AC   P01763;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig heavy chain V-III region WEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=83273707; PubMed=6410398;
RA   Goni F., Frangione B.;
RT   "Amino acid sequence of the Fv region of a human monoclonal IgM
RT   (protein WEA) with antibody activity against 3,4-pyruvylated
RT   galactose in Klebsiella polysaccharides K30 and K33.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).
```

```
CC   -!- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY
CC       AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH
CC       WALDENSTROM'S MACROGLOBULINEMIA.
DR   PIR; A02046; M3HUWE.
DR   HSSP; P01772; 21G2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin v region.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   NON_TER     114    114
SQ   SEQUENCE   114 AA;  12256 MW;  D88294FB418A07B7 CRC64;

Query Match             13.0%;  Score 6; DB 1; Length 114;
Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   23 IGGSGS 28
        ||||||
Db   51 IGGSGS 56

RESULT 13
HV3A_HUMAN
ID   HV3A_HUMAN     STANDARD;       PRT;    122 AA.
AC   P01762;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig heavy chain V-III region TRO.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE (MYELOMA PROTEIN TRO).
RX   MEDLINE=76023781; PubMed=809331;
RA   Kratzin H., Altevogt P., Ruban E., Kortt A., Staroscik K.,
RA   Hilschmann N.;
RT   "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),
RT   II. The amino acid sequence of the H-chain, alpha-type, subgroup III;
RT   structure of the complete IgA-molecule.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 356:1137-1342(1975).
CC   -!- MISCELLANEOUS: THE SEQUENCE OF THE C REGION IS ALSO GIVEN.
DR   PIR; A02045; AIHUTR.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin v region.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   NON_TER     122    122
SQ   SEQUENCE   122 AA;  13472 MW;  2E21A11DA04D80F9 CRC64;

Query Match             13.0%;  Score 6; DB 1; Length 122;
Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   23 IGGSGS 28
        ||||||
Db   51 IGGSGS 56

RESULT 14
RADC_COXBU
ID   RADC_COXBU     STANDARD;       PRT;    142 AA.
AC   O85403;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA repair protein radC (Fragment).
GN   RADC.
```

```
OS   Coxiella burnetii.
OC   Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC   Coxiella group; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NINE MILE PHASE I;
RA   Willems H., Jaeger C., Baljer G.;
RT   "Physical and genetic map of the obligate intracellular bacterium
RT   Coxiella burnetii.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RADC FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AF064960; AAD09944.1; -.
DR   InterPro; IPR001405; RadC.
DR   ProDom; PD007415; RadC; 1.
KW   DNA repair.
FT   NON_TER       1      1
SQ   SEQUENCE   142 AA;  15295 MW;  8C621E49EBFE4D43 CRC64;

Query Match             13.0%;  Score 6; DB 1; Length 142;
Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   21 IIIGGS 26
        ||||||
Db  119 IIIGGS 124

RESULT 15
Y805_AQUAE
ID   Y805_AQUAE     STANDARD;       PRT;    155 AA.
AC   O66989;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_805.
GN   AQ_805.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AE000708; AAC06954.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   155 AA;  18152 MW;  16933A2AC7344531 CRC64;
```

Query Match   13.0%; Score 6; DB 1; Length 155;
Best Local Similarity 100.0%; Pred. No. 38;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   35 LFNRSF 40
         ||||||
Db   54 LFNRSF 59

Search completed: November 5, 2002, 11:05:25
Job time : 57.3582 secs

OM protein - protein search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: November 5, 2002, 10:58:27 ; Search time 30.209 seconds
                                        (without alignments)
                                        263.424 Million cell updates/sec

Title: US-09-833-017B-2
Perfect score: 46
Sequence: 1 MKKTLSLKNDFKEIKTDELE.........GSLSTFFRLFNRSFTQALGK 46

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   SPTREMBL_19:*
            1: sp_archea:*
            2: sp_bacteria:*
            3: sp_fungi:*
            4: sp_human:*
            5: sp_invertebrate:*
            6: sp_mammal:*
            7: sp_mhc:*
            8: sp_organelle:*
            9: sp_phage:*
           10: sp_plant:*
           11: sp_rodent:*
           12: sp_virus:*
           13: sp_vertebrate:*
           14: sp_unclassified:*
           15: sp_rvirus:*
           16: sp_bacteriap:*
           17: sp_archeap:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	100.0	46	2	Q99QI5	streptococc
2	41	89.1	46	2	Q9APK7	streptococc
3	27	58.7	43	2	Q9APK6	streptococc
4	9	19.6	363		Q9R7H6	borrelia bu
5	7	15.2	88	16	Q98RG2	mycoplasma
6	7	15.2	96	12	Q83163	cauliflower
7	7	15.2	239	17	Q96Y40	sulfolobus
8	7	15.2	240	16	Q91488	pseudomonas
9	7	15.2	240	16	Q92WS3	rhizobium m
10	7	15.2	253	13	Q90315	coturnix co
11	7	15.2	253	16	O67894	aquifex aeo
12	7	15.2	268	16	Q98CC9	rhizobium l
13	7	15.2	276	10	Q9SEW3	glycine max
14	7	15.2	284	16	Q92S62	rhizobium m
15	7	15.2	287	13	Q90314	coturnix co
16	7	15.2	346	13	Q90WN4	xenopus lae
17	7	15.2	370	10	Q9LYK0	arabidopsis
18	7	15.2	411	4	Q14422	homo sapien
19	7	15.2	437	16	Q92EY2	listeria in
20	7	15.2	493	10	Q9LT85	arabidopsis
21	7	15.2	494	10	Q9FXA1	arabidopsis
22	7	15.2	579	4	Q9UEI7	homo sapien
23	7	15.2	612	12	Q9J5F6	fowlpox vir
24	7	15.2	618	10	O23121	arabidopsis
25	7	15.2	628	5	Q96OLO	drosophila
26	7	15.2	767	5	Q9YZ6	drosophila
27	7	15.2	808	16	Q9JVK8	neisseria m
28	7	15.2	929	5	Q9VR32	drosophila
29	7	15.2	1183	16	Q98QH9	mycoplasma
30	7	15.2	1358	5	O17892	caenorhabdi
31	7	15.2	1395	4	Q96SB7	homo sapien
32	7	15.2	2771	16	Q9VLS3	drosophila
33	7	15.2	3972	16	P73139	synechocyst
34	6	13.0	48	16	Q9PBI2	xylella fas
35	6	13.0	51	12	Q9XB45	escherichia
36	6	13.0	51	12	Q91G53	chilo iride
37	6	13.0	57	10	Q39398	betula verr
38	6	13.0	63	6	Q9XSB0	bos taurus
39	6	13.0	69	12	O41074	paramecium
40	6	13.0	69	12	Q91PP4	hepatitis b
41	6	13.0	97	10	Q9C7H0	arabidopsis
42	6	13.0	102	17	O58597	pyrococcus
43	6	13.0	104	6	Q9JY34	neisseria m
44	6	13.0	105	5	Q9U514	manduca sex
45	6	13.0	114	17	Q96Z48	sulfolobus
46	6	13.0	115	5	Q17259	brachionus
47	6	13.0	119	2	Q9RAF5	ralstonia s
48	6	13.0	120	2	Q9ALF6	uncultured
49	6	13.0	120	2	Q9ALF2	uncultured
50	6	13.0	120	2	Q9ALB0	uncultured
51	6	13.0	121	2	Q9ALF4	uncultured
52	6	13.0	121	2	Q9ALC3	uncultured
53	6	13.0	122	17	Q97AW7	thermoplasm
54	6	13.0	122	16	Q9RU75	deinococcus
55	6	13.0	127	16	Q9KUT9	vibrio chol
56	6	13.0	133	16	Q9A2R6	caulobacter
57	6	13.0	134	12	Q84484	paramecium
58	6	13.0	136	2	Q9RC70	bacillus ha
59	6	13.0	140	4	Q9BSC0	homo sapien
60	6	13.0	141	13	O57333	brachydanio
61	6	13.0	141	16	Q9JS86	chlamydia p
62	6	13.0	145	17	Q976P9	sulfolobus
63	6	13.0	152	5	Q08596	tenebrio mo
64	6	13.0	152	16	Q9DD2A9	mus musculu
65	6	13.0	154	16	Q9CHW9	lactococcus
66	6	13.0	156	11	Q99MF7	rattus norv
67	6	13.0	157	16	Q9JM24	listeria mo
68	6	13.0	167	2	Q9AKX7	legionella
69	6	13.0	167	10	Q9M6N1	cicer ariet
70	6	13.0	169	5	Q9NFN1	schistosoma
71	6	13.0	177	15	Q9QRM3	human immun
72	6	13.0	180	15	Q98FX0	rhizobium l
73	6	13.0	181	15	Q9O1F3	human immun
74	6	13.0	186	10	O04571	arabidopsis
75	6	13.0	186	11	Q62146	mus musculu
76	6	13.0	189	2	Q9KXW7	streptomyce
77	6	13.0	189	11	Q9CZV1	mus musculu
78	6	13.0	191	11	Q9D0T7	mus musculu
79	6	13.0	196	2	Q93KL7	uncultured
80	6	13.0	200	4	Q9BRI2	homo sapien
81	6	13.0	200	5	O16314	caenorhabdi
82	6	13.0	200	6	Q95LH6	bos taurus
83	6	13.0	201	13	Q83184	mink cell f
84	6	13.0	201	16	Q92R37	rhizobium m
85	6	13.0	202	15	Q9JBC7	human immun
86	6	13.0	202	16	Q9PLC4	chlamydia m
87	6	13.0	203	16	Q97R00	streptococc
88	6	13.0	204	11	Q9EPW1	rattus norv
89	6	13.0	205	5	O15772	trypanosoma

No.	n	thr	pos	id	Accession	Accession	Organism
90	6	13.0	205	5	O15777	O15777	trypanosoma
91	6	13.0	206	2	Q9AF58	Q9AF58	desulfotoma
92	6	13.0	206	2	Q9TR6	Q9TR6	uncultured
93	6	13.0	208	16	Q9CPJ3	Q9CPJ3	pasteurella
94	6	13.0	213	5	O60953	O60953	dictyosteli
95	6	13.0	213	16	O12993	O12993	oryzias lat
96	6	13.0	213	13	O84471	O84471	chlamydia t
97	6	13.0	215	13	O12999	O12999	xenopus lae
98	6	13.0	217	13	Q9PT27	Q9PT27	oncorhynchu
99	6	13.0	217	13	Q9PUS6	Q9PUS6	brachydanio
100	6	13.0	217	13	Q9DES0	Q9DES0	fugu rubrip
101	6	13.0	218	13	Q9PUS3	Q9PUS3	brachydanio
102	6	13.0	220	4	Q96RI4	Q96RI4	homo sapien
103	6	13.0	220	16	O25877	O25877	helicobacte
104	6	13.0	220	16	Q9ZJT8	Q9ZJT8	helicobacte
105	6	13.0	224	2	Q9X8G6	Q9X8G6	streptomyce
106	6	13.0	224	16	Q9V7H8	Q9V7H8	drosophila
107	6	13.0	224	5	Q9SSA6	Q9SSA6	drosophila
108	6	13.0	227	5	Q9X7J4	Q9X7J4	pseudomonas
109	6	13.0	227	16	Q51441	Q51441	pseudomonas
110	6	13.0	229	5	O62545	O62545	hydra atten
111	6	13.0	229	10	O41723	O41723	zinnia eleg
112	6	13.0	230	10	O13000	O13000	xenopus lae
113	6	13.0	232	10	Q98QM4	Q98QM4	mycoplasma
114	6	13.0	233	10	O23716	O23716	arabidopsis
115	6	13.0	234	17	Q9HSL9	Q9HSL9	halobacteri
116	6	13.0	235	3	Q9KR32	Q9KR32	vibrio chol
117	6	13.0	236	3	Q9P7V3	Q9P7V3	schizosacch
118	6	13.0	238	2	P94902	P94902	lactobacill
119	6	13.0	239	4	Q96J55	Q96J55	homo sapien
120	6	13.0	243	16	Q9Z8R7	Q9Z8R7	chlamydia p
121	6	13.0	243	16	Q930D1	Q930D1	rhizobium m
122	6	13.0	244	5	Q20804	Q20804	caenorhabdi
123	6	13.0	246	5	I18894	I18894	caenorhabdi
124	6	13.0	247	17	Q9LST9	Q9LST9	oryza sativ
125	6	13.0	248	17	Q97U1	Q97U1	thermoplasm
126	6	13.0	249	17	Q9HKF9	Q9HKF9	thermoplasm
127	6	13.0	250	16	Q9RB46	Q9RB46	clostridium
128	6	13.0	255	2	Q924F5	Q924F5	shigella so
129	6	13.0	256	4	Q96BD8	Q96BD8	homo sapien
130	6	13.0	256	12	O36455	O36455	chino del t
131	6	13.0	257	16	Q9A270	Q9A270	caulobacter
132	6	13.0	261	17	Q97BL7	Q97BL7	thermoplasm
133	6	13.0	262	17	Q9HL98	Q9HL98	thermoplasm
134	6	13.0	265	16	Q9PQE0	Q9PQE0	ureaplasma
135	6	13.0	270	16	O22782	O22782	arabidopsis
136	6	13.0	271	17	Q97W94	Q97W94	sulfolobus
137	6	13.0	271	4	Q9H027	Q9H027	homo sapien
138	6	13.0	271	16	Q9K8F9	Q9K8F9	bacillus ha
139	6	13.0	274	16	Q975C3	Q975C3	sulfolobus
140	6	13.0	276	16	Q9K9J7	Q9K9J7	bacillus ha
141	6	13.0	277	16	Q928Z8	Q928Z8	listeria in
142	6	13.0	278	5	Q9W5C7	Q9W5C7	drosophila
143	6	13.0	285	16	O66839	O66839	aquifex aeo
144	6	13.0	285	16	Q97JB3	Q97JB3	clostridium
145	6	13.0	287	5	Q94477	Q94477	dictyosteli
146	6	13.0	288	16	Q9KDP6	Q9KDP6	bacillus ha
147	6	13.0	289	16	Q97FG4	Q97FG4	clostridium
148	6	13.0	289	16	Q9PB74	Q9PB74	xylella fas
149	6	13.0	294	5	Q9V813	Q9V813	drosophila
150	6	13.0	294	5	O01509	O01509	caenorhabdi
151	6	13.0	297	4	Q93F53	Q93F53	pseudomonas
152	6	13.0	298	17	Q97286	Q97286	sulfolobus
153	6	13.0	300	11	Q64106	Q64106	rattus sp.
154	6	13.0	300	16	Q9KC62	Q9KC62	bacillus ha
155	6	13.0	300	2	Q9RJ57	Q9RJ57	streptomyce
156	6	13.0	304	16	Q9UN6	Q9UN6	vibrio vuln
157	6	13.0	305	17	Q9I259	Q9I259	pseudomonas
158	6	13.0	306	16	O58240	O58240	pyrococcus
159	6	13.0	306	16	Q9PPR5	Q9PPR5	ureaplasma
160	6	13.0	306	2	P71182	P71182	enterobacte
161	6	13.0	306	2	Q9AHH2	Q9AHH2	comamonas t
162	6	13.0	306	17	Q9V055	Q9V055	pyrococcus

No.	n	thr	pos	id	Accession	Accession	Organism
163	6	13.0	306	17	Q971C2	Q971C2	sulfolobus
164	6	13.0	308	11	Q9CVZ7	Q9CVZ7	mus musculu
165	6	13.0	309	17	Q9UYG8	Q9UYG8	pyrococcus
166	6	13.0	310	10	Q9ZTT0	Q9ZTT0	zea mays (m
167	6	13.0	312	5	O62365	O62365	caenorhabdi
168	6	13.0	313	16	Q98E28	Q98E28	rhizobium l
169	6	13.0	314	13	P94910	P94910	methanosarc
170	6	13.0	314	5	Q9L262	Q9L262	streptomyce
171	6	13.0	317	7	Q9VBH7	Q9VBH7	drosophila
172	6	13.0	318	16	Q93F56	Q93F56	pseudomonas
173	6	13.0	318	4	Q9UMO8	Q9UMO8	homo sapien
174	6	13.0	318	16	Q98707	Q98707	rhizobium l
175	6	13.0	319	5	O43950	O43950	theileria a
176	6	13.0	320	11	Q93F58	Q93F58	pseudomonas
177	6	13.0	321	17	Q923Q4	Q923Q4	mus musculu
178	6	13.0	322	17	Q97CK7	Q97CK7	thermoplasm
179	6	13.0	323	10	Q937X6	Q937X6	edwardsiell
180	6	13.0	323	16	Q98SD0	Q98SD0	guillardia
181	6	13.0	326	16	P74469	P74469	synechocyst
182	6	13.0	329	17	O83509	O83509	treponema p
183	6	13.0	330	3	Q9UYJ6	Q9UYJ6	pyrococcus
184	6	13.0	331	3	Q9HF69	Q9HF69	candida dub
185	6	13.0	336	3	O87519	O87519	escherichia
186	6	13.0	340	3	P78846	P78846	schizosacch
187	6	13.0	342	5	Q9Z4N6	Q9Z4N6	pasteurella
188	6	13.0	342	5	Q9VKR8	Q9VKR8	drosophila
189	6	13.0	345	4	Q9GRL0	Q9GRL0	leishmania
190	6	13.0	349	5	Q96IH2	Q96IH2	homo sapien
191	6	13.0	352	3	P91081	P91081	caenorhabdi
192	6	13.0	352	2	Q96W98	Q96W98	candida alb
193	6	13.0	352	2	Q93F60	Q93F60	pseudomonas
194	6	13.0	353	2	Q93F57	Q93F57	pseudomonas
195	6	13.0	354	16	Q93F48	Q93F48	pseudomonas
196	6	13.0	354	16	Q93F51	Q93F51	pseudomonas
197	6	13.0	354	16	Q93F52	Q93F52	pseudomonas
198	6	13.0	359	3	O25675	O25675	helicobacte
199	6	13.0	360	3	Q92M30	Q92M30	helicobacte
200	6	13.0	360	16	Q9V7Q2	Q9V7Q2	drosophila
201	6	13.0	360	16	Q00296	Q00296	beauveria b
202	6	13.0	366	16	Q9VXL2	Q9VXL2	drosophila
203	6	13.0	368	2	Q92816	Q92816	chlamydia p
204	6	13.0	369	2	Q99TT7	Q99TT7	staphylococ
205	6	13.0	370	16	Q93GD2	Q93GD2	desulfotoma
206	6	13.0	372	2	Q93GE2	Q93GE2	desulfotoma
207	6	13.0	372	2	Q95TX4	Q95TX4	drosophila
208	6	13.0	372	16	Q97PU9	Q97PU9	streptococc
209	6	13.0	372	2	Q57178	Q57178	salmonella
210	6	13.0	372	2	Q54034	Q54034	salmonella
211	6	13.0	372	2	Q54036	Q54036	salmonella
212	6	13.0	372	2	Q54025	Q54025	salmonella
213	6	13.0	372	2	Q54026	Q54026	salmonella
214	6	13.0	372	2	Q54027	Q54027	salmonella
215	6	13.0	372	2	Q54028	Q54028	salmonella
216	6	13.0	372	2	Q54029	Q54029	salmonella
217	6	13.0	372	2	Q54030	Q54030	salmonella
218	6	13.0	372	2	Q54031	Q54031	salmonella
219	6	13.0	372	2	Q54032	Q54032	salmonella
220	6	13.0	372	2	Q54037	Q54037	salmonella
221	6	13.0	372	2	Q54038	Q54038	salmonella
222	6	13.0	372	2	Q54039	Q54039	salmonella
223	6	13.0	372	2	Q54040	Q54040	salmonella
224	6	13.0	372	2	Q56052	Q56052	salmonella
225	6	13.0	375	10	Q93GE0	Q93GE0	desulfotoma
226	6	13.0	377	4	Q9SVA0	Q9SVA0	arabidopsis
227	6	13.0	377	4	Q96VM1	Q96VM1	beauveria b
228	6	13.0	378	4	Q96RQ1	Q96RQ1	homo sapien
229	6	13.0	379	3	Q9NZA3	Q9NZA3	homo sapien
230	6	13.0	379	3	Q02291	Q02291	fusarium sp
231	6	13.0	381	3	Q9Y759	Q9Y759	beauveria b
232	6	13.0	381	10	Q9XC30	Q9XC30	acinetobact
233	6	13.0	381	10	Q9CAQ3	Q9CAQ3	arabidopsis
234	6	13.0	386	6	Q9BGU5	Q9BGU5	bos taurus
235	6	13.0	387	10	Q9FN48	Q9FN48	arabidopsis

#				Code	Code	Description
236	13.0	6	5	Q9V313	Q9V313	drosophila
237	13.0	6	5	Q94JU8	Q94JU8	arabidopsis
238	13.0	6	10	Q93428	Q93428	chionodraco
239	13.0	6	13	Q9DEX3	Q9DEX3	clupea hare
240	13.0	6	13	Q00070	Q00070	aspergillus
241	13.0	6	13	P87370	P87370	oncorhynchu
242	13.0	6	3	Q9P8Y4	Q9P8Y4	coccidioide
243	13.0	6	13	Q9DD89	Q9DD89	brachydanio
244	13.0	6	13	Q9V717	Q9V717	drosophila
245	13.0	6	12	Q91FF5	Q91FF5	chilo iride
246	13.0	6	16	Q92EB0	Q92EB0	listeria in
247	13.0	6	1	O93650	O93650	archaeoglob
248	13.0	6	16	Q9PP65	Q9PP65	campylobact
249	13.0	6	12	Q9QD05	Q9QD05	tobacco vei
250	13.0	6	16	Q92WM5	Q92WM5	rhizobium m
251	13.0	6	5	O76607	O76607	caenorhabdi
252	13.0	6	5	Q9VKP7	Q9VKP7	drosophila
253	13.0	6	10	Q9MA53	Q9MA53	arabidopsis
254	13.0	6	16	Q9RUN5	Q9RUN5	deinococcus
255	13.0	6	3	Q06103	Q06103	saccharomyc
256	13.0	6	16	Q97GD5	Q97GD5	clostridium
257	13.0	6	16	Q97K95	Q97K95	clostridium
258	13.0	6	11	Q920C3	Q920C3	mus musculu
259	13.0	6	2	Q9SIA8	Q9SIA8	arabidopsis
260	13.0	6	2	Q93TD3	Q93TD3	pseudomonas
261	13.0	6	2	Q9UVS9	Q9UVS9	aspergillus
262	13.0	6	10	Q9LRP8	Q9LRP8	arabidopsis
263	13.0	6	8	Q35213	Q35213	oenothera b
264	13.0	6	16	Q9FI60	Q9FI60	corynebacte
265	13.0	6	10	Q9M158	Q9M158	arabidopsis
266	13.0	6	2	Q9X3V2	Q9X3V2	pseudomonas
267	13.0	6	16	Q9FHT1	Q9FHT1	arabidopsis
268	13.0	6	4	Q15846	Q15846	homo sapien
269	13.0	6	10	Q93VZ2	Q93VZ2	arabidopsis
270	13.0	6	16	Q92297	Q92297	streptococc
271	13.0	6	10	O82730	O82730	arabidopsis
272	13.0	6	3	Q9EZG5	Q9EZG5	streptococc
273	13.0	6	2	Q9UU82	Q9UU82	schizosacch
274	13.0	6	17	Q972D2	Q972D2	sulfolobus
275	13.0	6	2	P72120	P72120	pseudomonas
276	13.0	6	5	Q9W0H5	Q9W0H5	drosophila
277	13.0	6	17	O57804	O57804	pyrococcus
278	13.0	6	16	Q97SA7	Q97SA7	streptococc
279	13.0	6	17	Q9V2K8	Q9V2K8	pyrococcus
280	13.0	6	10	Q9JP99	Q9JP99	rhodocyclus
281	13.0	6	5	Q26475	Q26475	schistocerc
282	13.0	6	16	O06294	O06294	mycobacteri
283	13.0	6	2	Q9AD76	Q9AD76	streptomyce
284	13.0	6	4	Q9BXN7	Q9BXN7	homo sapien
285	13.0	6	3	Q9C4B1	Q9C4B1	aspergillus
286	13.0	6	3	Q12551	Q12551	aspergillus
287	13.0	6	5	Q9V764	Q9V764	drosophila
288	13.0	6	4	Q9H4D7	Q9H4D7	homo sapien
289	13.0	6	3	Q96X02	Q96X02	penicillium
290	13.0	6	3	Q96VA1	Q96VA1	aspergillus
291	13.0	6	3	Q96VA0	Q96VA0	aspergillus
292	13.0	6	5	Q96CL7	Q96CL7	homo sapien
293	13.0	6	5	Q20604	Q20604	caenorhabdi
294	13.0	6	11	Q920C2	Q920C2	mus musculu
295	13.0	6	1	O74288	O74288	emericella
296	13.0	6	10	Q942V7	Q942V7	oryza sativ
297	13.0	6	5	Q22017	Q22017	caenorhabdi
298	13.0	6	15	Q74230	Q74230	human immun
299	13.0	6	15	Q76633	Q76633	human immun
300	13.0	6	5	Q9V9Y4	Q9V9Y4	drosophila
301	13.0	6	10	Q9LWI6	Q9LWI6	oryza sativ
302	13.0	6	11	Q91V87	Q91V87	mus musculu
303	13.0	6	5	O87219	O87219	lactococcus
304	13.0	6	5	Q961B3	Q961B3	drosophila
305	13.0	6	2	Q68145	Q68145	burkholderi
306	13.0	6	2	Q93A25	Q93A25	buchnera ap
307	13.0	6	5	Q9VKZ2	Q9VKZ2	drosophila
308	13.0	6	5	Q95025	Q95025	dictyosteli

#				Code	Code	Description
309	13.0	6	4	Q9NVH1	Q9nvh1	homo sapien
310	13.0	6	16	Q25362	O25362	helicobacte
311	13.0	6	16	Q9ZLI9	Q9zli9	helicobacte
312	13.0	6	2	Q9EYF3	Q9eyf3	escherichia
313	13.0	6	5	O76134	O76134	tetrahymena
314	13.0	6	4	Q96FZ6	Q96fz6	homo sapien
315	13.0	6	2	Q9L900	Q9l900	enterococcu
316	13.0	6	10	Q64843	Q64843	arabidopsis
317	13.0	6	16	Q9PHX1	Q9phx1	campylobact
318	13.0	6	16	Q9I502	Q9i502	pseudomonas
319	13.0	6	5	O45066	O45066	caenorhabdi
320	13.0	6	10	Q9M056	Q9m056	arabidopsis
321	13.0	6	5	Q9V480	Q9v480	drosophila
322	13.0	6	5	O61711	O61711	caenorhabdi
323	13.0	6	5	Q27472	Q27472	caenorhabdi
324	13.0	6	11	Q9LK49	Q9lk49	arabidopsis
325	13.0	6	10	O88337	O88337	rattus norv
326	13.0	6	16	Q92HB9	Q92hb9	rickettsia
327	13.0	6	16	Q9ZCY6	Q9zcy6	rickettsia
328	13.0	6	16	P72416	P72416	streptococc
329	13.0	6	2	Q9JMY5	Q9jmy5	listeria mo
330	13.0	6	5	Q9SNZ0	Q9snz0	spodoptera
331	13.0	6	4	Q9NQX0	Q9nqx0	homo sapien
332	13.0	6	5	Q9VYA3	Q9vya3	drosophila
333	13.0	6	2	Q9US12	Q9us12	schizosacch
334	13.0	6	5	Q22779	Q22779	caenorhabdi
335	13.0	6	10	Q38957	Q38957	arabidopsis
336	13.0	6	16	Q94EX4	Q94ex4	cucumis mel
337	13.0	6	11	O88810	O88810	rattus norv
338	13.0	6	2	Q9JMY6	Q9jmy6	listeria mo
339	13.0	6	16	Q97MH1	Q97mh1	clostridium
340	13.0	6	4	Q969Y5	Q969y5	homo sapien
341	13.0	6	2	O33838	O33838	thermotoga
342	13.0	6	11	Q9CZ33	Q9cz33	mus musculu
343	13.0	6	16	Q9S5X2	Q9s5x2	thermotoga
344	13.0	6	5	Q9XZ22	Q9xz22	drosophila
345	13.0	6	17	Q980R1	Q980r1	sulfolobus
346	13.0	6	8	Q36521	Q36521	platymonas
347	13.0	6	5	Q9TZK7	Q9tzk7	caenorhabdi
348	13.0	6	11	Q9ROQ1	Q9roq1	mus musculu
349	13.0	6	16	Q983P2	Q983p2	rhizobium l
350	13.0	6	16	Q92BQ2	Q92bq2	listeria in
351	13.0	6	16	Q948P4	Q948p4	cucumis mel
352	13.0	6	11	Q9WV66	Q9wv66	mus musculu
353	13.0	6	17	Q96XH6	Q96xh6	sulfolobus
354	13.0	6	11	Q99NF0	Q99nf0	mus musculu
355	13.0	6	10	Q9FN17	Q9fn17	arabidopsis
356	13.0	6	17	Q28289	Q28289	archaeoglob
357	13.0	6	5	O02140	O02140	caenorhabdi
358	13.0	6	16	P95097	P95097	mycobacteri
359	13.0	6	16	Q9AAB8	Q9aab8	caulobacter
360	13.0	6	5	Q9EYG5	Q9eyg5	clostridium
361	13.0	6	8	Q94706	Q94706	physarum po
362	13.0	6	4	Q95560	Q95560	homo sapien
363	13.0	6	5	Q960E0	Q960e0	drosophila
364	13.0	6	5	Q9NSV0	Q9nsv0	caenorhabdi
365	13.0	6	4	Q9C0B6	Q9c0b6	homo sapien
366	13.0	6	5	Q9TVF4	Q9tvf4	caenorhabdi
367	13.0	6	13	Q42127	Q42127	xenopus lae
368	13.0	6	10	O13768	O13768	schizosacch
369	13.0	6	10	Q64770	Q64770	arabidopsis
370	13.0	6	5	Q62195	Q62195	caenorhabdi
371	13.0	6	5	Q9Y119	Q9y119	drosophila
372	13.0	6	5	Q9XTF3	Q9xtf3	caenorhabdi
373	13.0	6	5	Q9U434	Q9u434	drosophila
374	13.0	6	5	Q23484	Q23484	caenorhabdi
375	13.0	6	5	Q9VKR7	Q9vkr7	drosophila
376	13.0	6	5	Q44204	Q44204	agrobacteri
377	13.0	6	10	O64781	O64781	arabidopsis
378	13.0	6	2	Q9R6C4	Q9r6c4	agrobacteri
379	13.0	6	2	Q52297	Q52297	agrobacteri
380	13.0	6	5	Q9U3A8	Q9u3a8	caenorhabdi
381	13.0	6	10	Q94J7	Q94j7	oryza sativ

No.	n	Score	Pos	Len	Accession	Description
382	6	13.0	860		Q9NDT9	Q9ndt9 balanus amp
383	6	13.0	862		Q9W4P0	Q9w4p0 drosophila
384	6	13.0	872	4	Q9H6U7	Q9h6u7 homo sapien
385	6	13.0	872	4	Q96GN3	Q96gn3 homo sapien
386	6	13.0	874	10	O81723	O81723 arabidopsis
387	6	13.0	875	17	Q97WS1	Q97ws1 sulfolobus
388	6	13.0	879	5	O76978	O76978 sycon rapha
389	6	13.0	890	12	Q88171	Q88171 strawberry
390	6	13.0	893	16	Q9X1V4	Q9x1v4 thermotoga
391	6	13.0	909	6	Q9TTT4	Q9ttt4 oryctolagus
392	6	13.0	913		Q9ZFX7	Q9zfx7 pseudomonas
393	6	13.0	914		Q9R6Y9	Q9r6y9 alteromonas
394	6	13.0	945	11	Q91W15	Q91w15 mus musculu
395	6	13.0	955	17	Q974N4	Q974n4 sulfolobus
396	6	13.0	961	4	Q9Y2W1	Q9y2w1 homo sapien
397	6	13.0	968	10	Q9SXN6	Q9sxn6 nicotiana t
398	6	13.0	971	10	O04623	O04623 arabidopsis
399	6	13.0	977	5	Q9TYL9	Q9tyl9 caenorhabdi
400	6	13.0	977	10	Q9LLI0	Q9lli0 phaseolus l
401	6	13.0	978	16	O25308	O25308 helicobacte
402	6	13.0	1010	4	Q96PY8	Q96py8 homo sapien
403	6	13.0	1012	10	O15058	O15058 homo sapien
404	6	13.0	1013	10	O82677	O82677 chenopodium
405	6	13.0	1013	4	Q96PY9	Q96py9 homo sapien
406	6	13.0	1013	10	Q9LKZ3	Q9lkz3 arabidopsis
407	6	13.0	1021	10	Q93ZS5	Q93zs5 arabidopsis
408	6	13.0	1024	10	Q9LHH8	Q9lhh8 arabidopsis
409	6	13.0	1035	4	Q96RT8	Q96rt8 homo sapien
410	6	13.0	1054	5	Q21079	Q21079 caenorhabdi
411	6	13.0	1072	10	Q9C698	Q9c698 arabidopsis
412	6	13.0	1072	11	Q9DBT7	Q9dbt7 mus musculu
413	6	13.0	1086	16	Q92X83	Q92x83 rhizobium m
414	6	13.0	1088	5	Q9VM31	Q9vm31 drosophila
415	6	13.0	1110	10	Q9C7C9	Q9c7c9 arabidopsis
416	6	13.0	1120	13	Q91255	Q91255 petromyzon
417	6	13.0	1126	5	O17505	O17505 bombyx mori
418	6	13.0	1145	17	Q976A4	Q976a4 sulfolobus
419	6	13.0	1160	5	Q9IV56	Q9iv56 olive laten
420	6	13.0	1164	5	P90935	P90935 caenorhabdi
421	6	13.0	1174	16	Q9ZJN3	Q9zjn3 helicobacte
422	6	13.0	1177	5	Q9VB25	Q9vb25 drosophila
423	6	13.0	1184		O92611	O92611 pseudorabie
424	6	13.0	1218	10	O22725	O22725 arabidopsis
425	6	13.0	1222	5	Q9V5Q7	Q9v5q7 drosophila
426	6	13.0	1263	5	Q9V9J7	Q9v9j7 drosophila
427	6	13.0	1286	11	Q9QYZ2	Q9qyz2 mus musculu
428	6	13.0	1289		P90936	P90936 caenorhabdi
429	6	13.0	1291		Q9F0G7	Q9f0g7 rickettsia
430	6	13.0	1308		Q9UFV0	Q9ufv0 homo sapien
431	6	13.0	1308	12	Q9Q9Q8	Q9q9q8 soil-borne
432	6	13.0	1308	12	Q9Q9R0	Q9q9r0 soil-borne
433	6	13.0	1308	12	Q9Q906	Q9q906 soil-borne
434	6	13.0	1308	12	Q9QCE8	Q9qce8 soil-borne
435	6	13.0	1308	12	Q9DJG5	Q9djg5 soil-borne
436	6	13.0	1308	12	Q91DN2	Q91dn2 soil-borne
437	6	13.0	1320	12	O06359	O06359 soil-borne
438	6	13.0	1335		Q17250	Q17250 bombyx mori
439	6	13.0	1335	5	Q95PE2	Q95pe2 bombyx mori
440	6	13.0	1342	4	Q9VHH2	Q9vhh2 drosophila
441	6	13.0	1350	12	Q9WIE3	Q9wie3 chinese whe
442	6	13.0	1350	12	Q91QZ0	Q91qz0 chinese whe
443	6	13.0	1355	10	Q05700	Q05700 synechococc
444	6	13.0	1359	4	Q9UPR1	Q9upr1 homo sapien
445	6	13.0	1372	2	Q54151	Q54151 shigella fl
446	6	13.0	1373	2	Q9RAA2	Q9raa2 rickettsia
447	6	13.0	1373	2	Q9AL58	Q9al58 shigella fl
448	6	13.0	1429	10	Q94HU1	Q94hu1 oryza sativ
449	6	13.0	1475	5	Q9N4G4	Q9n4g4 caenorhabdi
450	6	13.0	1523		Q9N8U8	Q9n8u8 trypanosoma
451	6	13.0	1529	10	O81068	O81068 arabidopsis
452	6	13.0	1559	13	O73697	O73697 fugu rubrip
453	6	13.0	1583	10	Q9S7A7	Q9s7a7 oryza sativ
454	6	13.0	1602	11	Q91W25	Q91w25 mesocricetu
455	6	13.0	1643	11	Q91WX9	Q91wx9 rattus norv
456	6	13.0	1670	5	Q9U572	Q9u572 penaeus mon
457	6	13.0	1688	13	O57483	O57483 rana catesb
458	6	13.0	1769	4	Q9P273	Q9p273 homo sapien
459	6	13.0	1816	12	Q9JAD4	Q9jad4 soil-borne
460	6	13.0	1816	12	Q9JAD3	Q9jad3 soil-borne
461	6	13.0	1816	12	Q9JAD2	Q9jad2 soil-borne
462	6	13.0	1816	12	Q9IWB1	Q9iwb1 soil-borne
463	6	13.0	1816	12	Q9DJG6	Q9djg6 soil-borne
464	6	13.0	1816	12	Q91DN3	Q91dn3 soil-borne
465	6	13.0	1828	12	Q89249	Q89249 soil-borne
466	6	13.0	1855	5	Q9TX75	Q9tx75 plasmodium
467	6	13.0	1855	5	Q9BHN0	Q9bhn0 plasmodium
468	6	13.0	1858	12	Q9WJD7	Q9wjd7 chinese whe
469	6	13.0	1858	12	Q91QY9	Q91qy9 chinese whe
470	6	13.0	1864	10	Q9FYL7	Q9fyl7 arabidopsis
471	6	13.0	1876	5	Q95QQ8	Q95qq8 caenorhabdi
472	6	13.0	1877	5	Q9XXW1	Q9xxw1 plasmodium
473	6	13.0	1897	4	O13428	O13428 candida alb
474	6	13.0	1912	4	O95226	O95226 homo sapien
475	6	13.0	1912	5	Q9U0H1	Q9u0h1 plasmodium
476	6	13.0	1977	4	Q9UHB1	Q9uhb1 homo sapien
477	6	13.0	1981	11	Q92327	Q92327 rattus norv
478	6	13.0	1985	11	Q91IS7	Q91is7 mus musculu
479	6	13.0	2006	5	Q9VWF2	Q9vwf2 drosophila
480	6	13.0	2090	3	Q9P6X4	Q9p6x4 neurospora
481	6	13.0	2100	3	P87112	P87112 schizosacch
482	6	13.0	2110	5	Q9VRA6	Q9vra6 drosophila
483	6	13.0	2168	13	Q90ZA7	Q90za7 brachydanio
484	6	13.0	2169	11	Q9Z305	Q9z305 cavia porce
485	6	13.0	2271	16	Q99QY4	Q99qy4 staphylococ
486	6	13.0	2322	4	Q92675	Q92675 homo sapien
487	6	13.0	2346	11	Q91LC1	Q91lc1 mus musculu
488	6	13.0	2425	17	O28859	O28859 archaeoglob
489	6	13.0	2590	13	Q9W7R4	Q9w7r4 brachydanio
490	6	13.0	2693	13	Q91908	Q91908 xenopus lae
491	6	13.0	2706	5	Q97292	Q97292 plasmodium
492	6	13.0	2715	11	Q9WTS6	Q9wts6 mus musculu
493	6	13.0	2778	5	Q9V9T6	Q9v9t6 drosophila
494	6	13.0	3194	16	Q9ZLM3	Q9zlm3 helicobacte
495	6	13.0	3512	5	O62524	O62524 chironomus
496	6	13.0	3659	16	Q9BLN6	Q9bln6 rhizobium l
497	6	13.0	3734	16	Q9CIG0	Q9cig0 kallichroma
498	6	13.0	3930	16	Q98E20	Q98e20 rhizobium l
499	6	13.0	4910	3	Q12019	Q12019 saccharomyc
500	6	13.0	8243	6	O96554	O96554 cryptospori
501	6	13.0	13288	6	O18758	O18758 sus scrofa
502	6	13.0	26926	6	Q10466	Q10466 homo sapien
503	5	10.9	14	4	Q9O061	Q9o061 bos taurus
504	5	10.9	16	11	O35917	O35917 rattus norv
505	5	10.9	25	5	Q9R4L0	Q9r4l0 spiroplasma
506	5	10.9	26	9	Q9TRV9	Q9trv9 oryctolagus
507	5	10.9	29	16	Q38370	Q38370 mycobacteri
508	5	10.9	36	11	Q9JJY5	Q9jy5 mus musculu
509	5	10.9	40	16	O50887	O50887 borrelia bu
510	5	10.9	40	6	Q9DDE8	Q9dde8 sus scrofa
511	5	10.9	41	6	Q91FE0	Q9fe0 chilo iride
512	5	10.9	42	2	Q93LX0	Q9lx0 helicobacte
513	5	10.9	42	2	P70476	P70476 rattus norv
514	5	10.9	46	11	Q98Q40	Q98q40 mycoplasma
515	5	10.9	46	16	Q51952	Q51952 staphylococ
516	5	10.9	46	2	Q15456	Q15456 homo sapien
517	5	10.9	47	12	O71201	O71201 bovine aden
518	5	10.9	48	16	Q9JS22	Q9js22 neisseria m
519	5	10.9	49	16	O25662	O25662 helicobacte
520	5	10.9	49	2	Q9F214	Q9f214 roseateles
521	5	10.9	49	2	Q9FB90	Q9fb90 helicobacte
522	5	10.9	49	2	Q9FB89	Q9fb89 helicobacte
523	5	10.9	49	2	Q9FB85	Q9fb85 helicobacte
524	5	10.9	49	2	Q9FB84	Q9fb84 helicobacte
525	5	10.9	49	2	Q9FB83	Q9fb83 helicobacte
526	5	10.9	50	13	Q9DDZ2	Q9dz2 micropterus
527	5	10.9	50	16	O50723	O50723 borrelia bu

us-09-833-017b-2.oligo.rspt

Index	n1	n2	ID	Description	ID2	10.9	5
528	51	16	O51018	borrelia bu	O51018	10.9	5
529	51	16	Q9RNP7	vibrio chol	Q9RNP7	10.9	5
530	52	6	O77483	canis famil	O77483	10.9	5
531	52	16	Q9PDF0	xylella fas	Q9PDF0	10.9	5
532	53	2	Q93SL5	streptococc	Q93SL5	10.9	5
533	53	16	O65063	picea maria	O65063	10.9	5
534	53	16	Q9PD01	xylella fas	Q9PD01	10.9	5
535	53	16	O69194	streptococc	O69194	10.9	5
536	54	6	Q23882	dictyosteli	Q23882	10.9	5
537	54	6	Q28539	ovis aries	Q28539	10.9	5
538	55	8	Q9XDS5	salmonella	Q9XDS5	10.9	5
539	55	8	Q9TBJ4	cacomantis	Q9TBJ4	10.9	5
540	56	13	O57575	cynops pyrr	O57575	10.9	5
541	56	16	Q92RV4	rhizobium m	Q92RV4	10.9	5
542	57	2	O52965	enterococcu	O52965	10.9	5
543	57	4	Q96PS3	homo sapien	Q96PS3	10.9	5
544	58	5	Q95SA9	drosophila	Q95SA9	10.9	5
545	58	16	Q97JS7	clostridium	Q97JS7	10.9	5
546	60	10	Q41933	arabidopsis	Q41933	10.9	5
547	61	16	Q23877	dictyosteli	Q23877	10.9	5
548	61	6	Q26829	trypanosoma	Q26829	10.9	5
549	62	16	P76618	escherichia	P76618	10.9	5
550	63	9	O64074	bacteriopha	O64074	10.9	5
551	63	16	O31947	bacillus su	O31947	10.9	5
552	64	6	Q9F1E0	acetobacter	Q9F1E0	10.9	5
553	64	10	Q9ANN8	bradyrhizob	Q9ANN8	10.9	5
554	64	10	Q9LNS5	arabidopsis	Q9LNS5	10.9	5
555	65	16	Q23878	dictyosteli	Q23878	10.9	5
556	66	8	Q23609	caenorhabdi	Q23609	10.9	5
557	67	8	O20149	chlorella v	O20149	10.9	5
558	67	11	Q9ER13	rattus norv	Q9ER13	10.9	5
559	67	16	Q9JZM3	neisseria m	Q9JZM3	10.9	5
560	67	16	Q9JY08	neisseria m	Q9JY08	10.9	5
561	67	16	Q9CF38	lactococcus	Q9CF38	10.9	5
562	68	7	O19350	equus cabal	O19350	10.9	5
563	69	12	O48131	phrynosoma	O48131	10.9	5
564	69	12	Q83161	cauliflower	Q83161	10.9	5
565	70	4	Q9H4S1	homo sapien	Q9H4S1	10.9	5
566	70	16	Q98H86	rhizobium l	Q98H86	10.9	5
567	71	9	Q38564	bacteriopha	Q38564	10.9	5
568	71	16	Q9JI77	mus musculu	Q9JI77	10.9	5
569	71	17	Q978M8	thermoplasm	Q978M8	10.9	5
570	72	2	Q44537	azotobacter	Q44537	10.9	5
571	72	5	Q9TW80	drosophila	Q9TW80	10.9	5
572	72	11	Q9NI13	plasmodium	Q9NI13	10.9	5
573	72	11	Q61682	mus musculu	Q61682	10.9	5
574	72	16	P74009	synechocyst	P74009	10.9	5
575	74	2	O32312	bacillus th	O32312	10.9	5
576	74	12	Q95UY3	plasmodium	Q95UY3	10.9	5
577	74	12	Q9DW87	rat cytomeg	Q9DW87	10.9	5
578	75	11	Q9WUN3	rattus norv	Q9WUN3	10.9	5
579	75	11	Q9QPM9	human immun	Q9QPM9	10.9	5
580	75	15	Q9QPM8	human immun	Q9QPM8	10.9	5
581	75	15	Q9QPM7	human immun	Q9QPM7	10.9	5
582	75	15	Q9QPM5	human immun	Q9QPM5	10.9	5
583	75	15	Q9QPM3	human immun	Q9QPM3	10.9	5
584	75	15	Q9QPM2	human immun	Q9QPM2	10.9	5
585	75	15	Q9QPM1	human immun	Q9QPM1	10.9	5
586	75	15	Q9QPM0	human immun	Q9QPM0	10.9	5
587	75	15	Q9QPL9	human immun	Q9QPL9	10.9	5
588	75	15	Q9QPL8	human immun	Q9QPL8	10.9	5
589	75	15	Q9QPL7	human immun	Q9QPL7	10.9	5
590	75	15	Q9QPL6	human immun	Q9QPL6	10.9	5
591	75	15	Q9QPL5	human immun	Q9QPL5	10.9	5
592	75	15	Q9QPL4	human immun	Q9QPL4	10.9	5
593	75	15	Q9QPL3	human immun	Q9QPL3	10.9	5
594	75	15	Q9QPL2	human immun	Q9QPL2	10.9	5
595	75	15	Q9QPL1	human immun	Q9QPL1	10.9	5
596	75	15	Q9QPL0	human immun	Q9QPL0	10.9	5
597	75	15	Q9QPK9	human immun	Q9QPK9	10.9	5
598	75	15	Q9QPK8	human immun	Q9QPK8	10.9	5
599	75	15	Q9QPK7	human immun	Q9QPK7	10.9	5
600	75	15	Q9QPK6	human immun	Q9QPK6	10.9	5

Index	ID	Description	ID2	15	75	10.9	5
601	Q9qpk4	human immun	Q9QPK4	15	75	10.9	5
602	Q9qpk3	human immun	Q9QPK3	15	75	10.9	5
603	Q9qpk2	human immun	Q9QPK2	15	75	10.9	5
604	Q9qpk1	human immun	Q9QPK1	15	75	10.9	5
605	Q9qpk0	human immun	Q9QPK0	15	75	10.9	5
606	Q9qpj9	human immun	Q9QPJ9	15	75	10.9	5
607	Q9qpj7	human immun	Q9QPJ7	15	75	10.9	5
608	Q9qpj6	human immun	Q9QPJ6	15	75	10.9	5
609	Q9qpj5	human immun	Q9QPJ5	15	75	10.9	5
610	Q9qpj4	human immun	Q9QPJ4	15	75	10.9	5
611	Q9qpj3	human immun	Q9QPJ3	15	75	10.9	5
612	Q9qpj2	human immun	Q9QPJ2	15	75	10.9	5
613	Q9qpj0	human immun	Q9QPJ0	15	75	10.9	5
614	Q9qpi9	human immun	Q9QPI9	15	75	10.9	5
615	Q9qpi8	human immun	Q9QPI8	15	75	10.9	5
616	Q9qpi7	human immun	Q9QPI7	15	75	10.9	5
617	Q9qpi6	human immun	Q9QPI6	15	75	10.9	5
618	Q9qpi5	human immun	Q9QPI5	15	75	10.9	5
619	Q9qpi4	human immun	Q9QPI4	15	75	10.9	5
620	Q9qpi3	human immun	Q9QPI3	15	75	10.9	5
621	Q9qpi2	human immun	Q9QPI2	15	75	10.9	5
622	Q9qpi1	human immun	Q9QPI1	15	75	10.9	5
623	Q9qpi0	human immun	Q9QPI0	15	75	10.9	5
624	Q9qph9	human immun	Q9QPH9	15	75	10.9	5
625	Q9qph8	human immun	Q9QPH8	15	75	10.9	5
626	Q9qph7	human immun	Q9QPH7	15	75	10.9	5
627	Q9qph6	human immun	Q9QPH6	15	75	10.9	5
628	Q9qph5	human immun	Q9QPH5	15	75	10.9	5
629	Q9qph4	human immun	Q9QPH4	15	75	10.9	5
630	Q9qph3	human immun	Q9QPH3	15	75	10.9	5
631	Q9qph2	human immun	Q9QPH2	15	75	10.9	5
632	Q9qph1	human immun	Q9QPH1	15	75	10.9	5
633	Q9qph0	human immun	Q9QPH0	15	75	10.9	5
634	Q9qpg9	human immun	Q9QPG9	15	75	10.9	5
635	Q9qpg8	human immun	Q9QPG8	15	75	10.9	5
636	Q9qpg7	human immun	Q9QPG7	15	75	10.9	5
637	Q9qpg6	human immun	Q9QPG6	15	75	10.9	5
638	Q9qpg4	human immun	Q9QPG4	15	75	10.9	5
639	Q9qpg3	human immun	Q9QPG3	15	75	10.9	5
640	Q9qpg2	human immun	Q9QPG2	15	75	10.9	5
641	Q9qpg1	human immun	Q9QPG1	15	75	10.9	5
642	Q9qpg0	human immun	Q9QPG0	15	75	10.9	5
643	Q9qpf9	human immun	Q9QPF9	15	75	10.9	5
644	Q9qpf8	human immun	Q9QPF8	15	75	10.9	5
645	Q9qpf7	human immun	Q9QPF7	15	75	10.9	5
646	Q9qpf6	human immun	Q9QPF6	15	75	10.9	5
647	Q9qpf5	human immun	Q9QPF5	15	75	10.9	5
648	Q9qpf4	human immun	Q9QPF4	15	75	10.9	5
649	Q9qpf3	human immun	Q9QPF3	15	75	10.9	5
650	Q9qpf2	human immun	Q9QPF2	15	75	10.9	5
651	Q9qpf0	human immun	Q9QPF0	15	75	10.9	5
652	Q9qpe9	human immun	Q9QPE9	15	75	10.9	5
653	Q9qpe8	human immun	Q9QPE8	15	75	10.9	5
654	Q9qpe7	human immun	Q9QPE7	15	75	10.9	5
655	Q9qpe6	human immun	Q9QPE6	15	75	10.9	5
656	Q9qpe5	human immun	Q9QPE5	15	75	10.9	5
657	Q9qpe4	human immun	Q9QPE4	15	75	10.9	5
658	Q9qpe3	human immun	Q9QPE3	15	75	10.9	5
659	Q9qpe2	human immun	Q9QPE2	15	75	10.9	5
660	Q9qpe1	human immun	Q9QPE1	15	75	10.9	5
661	Q9qpe0	human immun	Q9QPE0	15	75	10.9	5
662	Q9qpd8	human immun	Q9QPD8	15	75	10.9	5
663	Q9qpd7	human immun	Q9QPD7	15	75	10.9	5
664	Q9qpd5	human immun	Q9QPD5	15	75	10.9	5
665	Q9qpd4	human immun	Q9QPD4	15	75	10.9	5
666	Q9qpd3	human immun	Q9QPD3	15	75	10.9	5
667	Q9qpd2	human immun	Q9QPD2	15	75	10.9	5
668	Q9qpd1	human immun	Q9QPD1	15	75	10.9	5
669	Q9pnd7	campylobact	Q9PND7	16	76	10.9	5
670	o40946	pharbitis n	O40946	10	76	10.9	5
671	o40947	pharbitis n	O40947	10	76	10.9	5
672	o40952	pharbitis n	O40952	10	76	10.9	5
673	Q9rrn2	deinococcus	Q9RRN2	16	76	10.9	5

#			Accession			Description
674	10.9	5	Q92GZ1	77	2	Q92gz1 yersinia pe
675	10.9	5	Q93MD7	77	2	Q93md7 clostridium
676	10.9	5	O08631	77	11	O08631 mus musculu
677	10.9	5	Q9ZHK9	78	2	Q9zhk9 haemophilus
678	10.9	5	Q9YP99	78	12	Q9yp99 canine herp
679	10.9	5	Q92N05	78	16	Q9zn05 helicobacte
680	10.9	5	Q92U06	79	11	Q9zu6 mus musculu
681	10.9	5	Q9PDY5	79	16	Q9pdy5 xylella fas
682	10.9	5	Q9VN17	80	5	Q9vn17 drosophila
683	10.9	5	O23483	80	10	O23483 arabidopsis
684	10.9	5	Q9RON2	81	11	Q9r0n2 rattus norv
685	10.9	5	Q97HR0	81	16	Q97hr0 clostridium
686	10.9	5	Q93MT9	82	2	Q93mt9 leptospira
687	10.9	5	Q69144	82	12	Q69144 human herpe
688	10.9	5	O41082	82	12	O41082 paramecium
689	10.9	5	Q98GZ6	82	16	Q98gz6 rhizobium l
690	10.9	5	Q9BMC1	83	5	Q9bmc1 ceratitis c
691	10.9	5	Q23875	83	5	Q23875 dictyosteli
692	10.9	5	Q97M00	83	16	Q97m00 clostridium
693	10.9	5	Q92Q95	83	16	Q92q95 rhizobium m
694	10.9	5	Q9K8T4	84	16	Q9k8t4 bacillus ha
695	10.9	5	Q984M9	84	16	Q984m9 rhizobium l
696	10.9	5	Q97UP8	84	17	Q97up8 sulfolobus
697	10.9	5	Q49461	85	5	Q49461 mycoplasma
698	10.9	5	Q9UE33	85	4	Q9ue33 homo sapien
699	10.9	5	Q94364	85	5	Q94364 caenorhabdi
700	10.9	5	Q98MY3	85	16	Q98my3 rhizobium l
701	10.9	5	Q9B8X2	86	8	Q9b8x2 schistosoma
702	10.9	5	Q95U07	86	8	Q958u7 echinococcu
703	10.9	5	Q920Y5	86	11	Q920y5 mus musculu
704	10.9	5	Q9AIF2	87	2	Q9aif2 candidatus
705	10.9	5	Q23879	87	5	Q23879 dictyosteli
706	10.9	5	Q94KC4	87	10	Q94kc4 arabidopsis
707	10.9	5	Q97EY0	87	16	Q97ey0 clostridium
708	10.9	5	Q9RQH4	88	5	Q9rqh4 listeria mo
709	10.9	5	Q23880	88	5	Q23880 dictyosteli
710	10.9	5	Q97588	88	6	Q97588 oryctolagus
711	10.9	5	Q9LI99	88	10	Q9li99 arabidopsis
712	10.9	5	Q93MI6	89	2	Q93mi6 azotobacter
713	10.9	5	Q95S05	89	5	Q95s05 drosophila
714	10.9	5	Q90VX5	89	13	Q90vx5 pseudopleur
715	10.9	5	Q9HWZ4	89	16	Q9hwz4 pseudomonas
716	10.9	5	Q98A25	89	16	Q98a25 rhizobium l
717	10.9	5	Q9NI24	90	5	Q9ni24 plasmodium
718	10.9	5	Q9JFL6	90	15	Q9jfl6 human immun
719	10.9	5	Q9B0L1	90	16	Q9b0l1 mycoplasma
720	10.9	5	Q977A4	90	17	Q977a4 sulfolobus
721	10.9	5	Q28857	91	6	Q28857 pan troglod
722	10.9	5	O02185	91	12	O02185 choristoneu
723	10.9	5	Q91LQ2	91	16	Q921q2 rhizobium m
724	10.9	5	Q01695	92	5	Q01695 penicillium
725	10.9	5	P82116	92	5	P82116 drosophila
726	10.9	5	Q9WBE3	92	15	Q9wbe3 human immun
727	10.9	5	Q9K2M8	92	16	Q9k2m8 vibrio chol
728	10.9	5	Q9JR94	92	16	Q9jr94 neisseria m
729	10.9	5	Q55253	93	9	Q55253 streptococc
730	10.9	5	Q37978	93	9	Q37978 bacteriopha
731	10.9	5	Q92FH9	93	16	Q92fh9 listeria in
732	10.9	5	Q56942	94	2	Q56942 yersinia pe
733	10.9	5	Q48582	95	2	Q48582 listeria in
734	10.9	5	Q00953	95	3	Q00953 kluyveromyc
735	10.9	5	Q9W9W3	95	15	Q9w9w3 human immun
736	10.9	5	Q9W9N5	95	15	Q9w9n5 human immun
737	10.9	5	Q77320	95	17	Q77320 human immun
738	10.9	5	O28676	96	17	O28676 archaeoglob
739	10.9	5	Q9F6C4	96	2	Q9f6c4 propionibac
740	10.9	5	Q93VC1	96	10	Q93vc1 oryza sativ
741	10.9	5	Q83180	96	12	Q83180 cauliflower
742	10.9	5	Q9WI31	96	12	Q9wi31 cauliflower
743	10.9	5	Q90X51	96	13	Q90x51 fundulus he
744	10.9	5	Q42187	97	10	Q42187 arabidopsis
745	10.9	5	Q9EBG1	97	12	Q9ebg1 porcine ade
746	10.9	5	Q9K5P2	97	16	Q9k5p2 bacillus ha
747	10.9	5	Q97S59	97	16	Q97s59 streptococc
748	10.9	5	Q48467	98	9	Q48467 bacteriopha
749	10.9	5	Q9ZTI9	98	10	Q9zti9 zea mays (m
750	10.9	5	Q9J8B4	98	12	Q9j8b4 spodoptera
751	10.9	5	Q9YWS7	98	15	Q9yws7 human immun
752	10.9	5	Q99YF9	98	15	Q99yf9 streptococc
753	10.9	5	Q9AFU5	99	2	Q9afu5 shigella fl
754	10.9	5	Q9U771	99	5	Q9u771 plasmodium
755	10.9	5	O10115	99	15	O10115 human immun
756	10.9	5	Q9CBF0	99	16	Q9cbf0 mycobacteri
757	10.9	5	Q92T24	99	16	Q92t24 rhizobium m
758	10.9	5	Q9XYF3	100	5	Q9xyf3 drosophila
759	10.9	5	Q9XYF2	100	5	Q9xyf2 drosophila
760	10.9	5	Q9XYF1	100	5	Q9xyf1 drosophila
761	10.9	5	Q9XYF0	100	5	Q9xyf0 drosophila
762	10.9	5	Q9XYE9	100	5	Q9xye9 drosophila
763	10.9	5	Q9XYE8	100	5	Q9xye8 drosophila
764	10.9	5	Q9XYE7	100	5	Q9xye7 drosophila
765	10.9	5	Q9XYE6	100	5	Q9xye6 drosophila
766	10.9	5	Q9XYE5	100	5	Q9xye5 drosophila
767	10.9	5	Q9XYE4	100	5	Q9xye4 drosophila
768	10.9	5	Q9XYE3	100	5	Q9xye3 drosophila
769	10.9	5	Q9XYE2	100	5	Q9xye2 drosophila
770	10.9	5	Q9XYE1	100	5	Q9xye1 drosophila
771	10.9	5	Q9XYE0	100	5	Q9xye0 drosophila
772	10.9	5	Q9TYD9	100	5	Q9tyd9 styela clav
773	10.9	5	Q93X17	100	10	Q93x17 pharbitis n
774	10.9	5	Q9A7J8	100	16	Q9a7j8 caulobacter
775	10.9	5	O88540	100	16	O88540 rattus norv
776	10.9	5	Q9PPB2	101	16	Q9ppb2 campylobact
777	10.9	5	Q9KFB0	102	16	Q9kfb0 bacillus ha
778	10.9	5	Q928K3	102	16	Q928k3 listeria in
779	10.9	5	Q05457	103	3	Q05457 saccharomyc
780	10.9	5	Q9NLT9	103	5	Q9nlt9 leishmania
781	10.9	5	O15941	103	5	O15941 bombyx mori
782	10.9	5	Q9ZRW1	103	10	Q9zrw1 cicer ariet
783	10.9	5	Q9E8F7	103	12	Q9e8f7 porcine ade
784	10.9	5	Q71731	103	16	Q71731 human immun
785	10.9	5	Q9JSA8	103	16	Q9jsa8 chlamydia p
786	10.9	5	Q9MFC3	104	8	Q9mfc3 beta vulgar
787	10.9	5	Q9AZH1	104	9	Q9azh1 bacteriopha
788	10.9	5	Q39084	104	16	Q39084 arabidopsis
789	10.9	5	Q9CEA3	104	16	Q9cea3 lactococcus
790	10.9	5	Q99ZL7	104	16	Q99zl7 streptococc
791	10.9	5	Q98I11	104	17	Q98i11 rhizobium l
792	10.9	5	O57995	104	17	O57995 pyrococcus
793	10.9	5	Q98524	105	12	Q98524 paramecium
794	10.9	5	Q9PQR7	105	16	Q9pqr7 ureaplasma
795	10.9	5	Q9JWB2	105	16	Q9jwb2 neisseria m
796	10.9	5	Q9MD41	106	8	Q9md41 saccharomyc
797	10.9	5	O50937	106	16	O50937 borrelia bu
798	10.9	5	Q9A169	106	16	Q9a169 streptococc
799	10.9	5	Q92GI8	106	16	Q92gi8 rickettsia
800	10.9	5	Q9M0B5	107	10	Q9m0b5 arabidopsis
801	10.9	5	Q9D5B6	107	11	Q9d5b6 mus musculu
802	10.9	5	O75678	107	15	O75678 human immun
803	10.9	5	O75683	107	15	O75683 human immun
804	10.9	5	Q75684	107	15	Q75684 human immun
805	10.9	5	Q75685	107	15	Q75685 human immun
806	10.9	5	O29702	107	17	O29702 archaeoglob
807	10.9	5	O73985	107	17	O73985 pyrococcus
808	10.9	5	Q9F3G7	108	2	Q9f3g7 streptomyce
809	10.9	5	Q934L7	108	5	Q934l7 vibrio salm
810	10.9	5	Q9VB37	108	5	Q9vb37 drosophila
811	10.9	5	O02562	108	5	O02562 acrolepiops
812	10.9	5	Q9XST8	108	6	Q9xst8 canis famil
813	10.9	5	Q9SXR1	108	10	Q9sxr1 oryza sativ
814	10.9	5	Q940A0	108	10	Q940a0 arabidopsis
815	10.9	5	O32892	108	16	O32892 mycobacteri
816	10.9	5	Q9L7X0	109	2	Q9l7x0 clostridium
817	10.9	5	Q9ANF1	109	2	Q9anf1 bradyrhizob
818	10.9	5	Q9X7N0	109	2	Q9x7n0 streptomyce
819	10.9	5	P91881	109	5	P91881 diadromus p

No.	Len	Cnt	Accession	Description		(5)	(10.9)
820	109	5	P91877	P91877 diprion pin	5	5	10.9
821	109	10	Q9FV31	Q9fv31 pisum sativ		5	10.9
822	110	10	Q38474	Q38474 bacteriopha		5	10.9
823	110	11	Q9D3T2	Q9d3t2 mus musculu		5	10.9
824	110	11	Q61977	Q61977 mus musculu		5	10.9
825	110	16	Q931X8	Q931x8 staphylococ		5	10.9
826	111	5	Q9NI26	Q9ni26 plasmodium		5	10.9
827	111	16	Q9D9Y7	Q9d9y7 mus musculu		5	10.9
828	111	16	Q923A0	Q923a0 escherichia		5	10.9
829	112	5	Q9XVH1	Q9xvh1 caenorhabdi		5	10.9
830	112	3	Q00745	Q00745 emericella		5	10.9
831	113	12	Q89271	Q89271 carnation m		5	10.9
832	113	13	Q91153	Q91153 notophthalm		5	10.9
833	113	13	Q9KFU4	Q9kfu4 bacillus ha		5	10.9
834	114	10	Q41211	Q41211 striga asia		5	10.9
835	114	17	O58083	O58083 pyrococcus		5	10.9
836	115	3	O59983	O59983 aspergillus		5	10.9
837	115	3	O59988	O59988 aspergillus		5	10.9
838	115	3	O59989	O59989 aspergillus		5	10.9
839	115	3	O59990	O59990 neosartorya		5	10.9
840	115	3	O59991	O59991 neosartorya		5	10.9
841	115	3	O59994	O59994 neosartorya		5	10.9
842	115	3	O59995	O59995 neosartorya		5	10.9
843	115	3	O59997	O59997 neosartorya		5	10.9
844	115	3	O60001	O60001 neosartorya		5	10.9
845	115	3	O60002	O60002 neosartorya		5	10.9
846	115	3	Q19616	Q19616 caenorhabdi		5	10.9
847	115	9	O64164	O64164 bacteriopha		5	10.9
848	115	12	Q9I161	Q9i161 regina rana		5	10.9
849	115	16	O31884	O31884 bacillus su		5	10.9
850	115	16	Q9CFJ9	Q9cfj9 lactococcus		5	10.9
851	116	2	Q9EUH8	Q9euh8 salmonella		5	10.9
852	116	8	Q9BJQ2	Q9bjq2 plasmodium		5	10.9
853	116	8	Q9B8A4	Q9b8a4 trichinella		5	10.9
854	116	13	Q9W787	Q9w787 ctenophorus		5	10.9
855	116	16	Q9X1V2	Q9x1v2 thermotoga		5	10.9
856	116	17	O27245	O27245 methanother		5	10.9
857	117	5	Q9BJP9	Q9bjp9 plasmodium		5	10.9
858	117	10	Q9SXR0	Q9sxr0 oryza sativ		5	10.9
859	117	10	Q9SSS6	Q9sss6 arabidopsis		5	10.9
860	117	17	Q9YC28	Q9yc28 aeropyrum p		5	10.9
861	118	10	Q9RNT8	Q9rnt8 solar lake		5	10.9
862	118	10	Q9T0H7	Q9t0h7 arabidopsis		5	10.9
863	118	10	Q9FYI2	Q9fyi2 arabidopsis		5	10.9
864	118	12	Q949R3	Q949r3 arabidopsis		5	10.9
865	118	12	Q9YW07	Q9yw07 melanoplus		5	10.9
866	118	12	Q9YMG9	Q9ymg9 lymantria d		5	10.9
867	118	16	Q988K6	Q988k6 rhizobium l		5	10.9
868	118	16	Q92KJ5	Q92kj5 rhizobium m		5	10.9
869	119	2	Q9XD86	Q9xd86 corynebacte		5	10.9
870	119	2	Q9ALE4	Q9ale4 uncultured		5	10.9
871	119	2	Q9ALE2	Q9ale2 uncultured		5	10.9
872	119	2	Q9ALE1	Q9ale1 uncultured		5	10.9
873	119	2	Q9ALC9	Q9alc9 uncultured		5	10.9
874	119	5	Q9ALC8	Q9alc8 uncultured		5	10.9
875	119	5	Q9GQZ0	Q9gqz0 plasmodium		5	10.9
876	119	10	Q9T0L7	Q9t0l7 arabidopsis		5	10.9
877	119	16	O67530	O67530 aquifex aeo		5	10.9
878	119	17	Q980M9	Q980m9 sulfolobus		5	10.9
879	120	2	Q9RNT9	Q9rnt9 solar lake		5	10.9
880	120	2	Q9RNT5	Q9rnt5 solar lake		5	10.9
881	120	2	Q9RNT2	Q9rnt2 solar lake		5	10.9
882	120	2	Q9RNT0	Q9rnt0 solar lake		5	10.9
883	120	2	Q9ALH9	Q9alh9 uncultured		5	10.9
884	120	2	Q9ALH7	Q9alh7 uncultured		5	10.9
885	120	2	Q9ALH6	Q9alh6 uncultured		5	10.9
886	120	2	Q9ALH4	Q9alh4 uncultured		5	10.9
887	120	2	Q9ALH3	Q9alh3 uncultured		5	10.9
888	120	2	Q9ALH2	Q9alh2 uncultured		5	10.9
889	120	2	Q9ALH1	Q9alh1 uncultured		5	10.9
890	120	2	Q9ALH0	Q9alh0 uncultured		5	10.9
891	120	2	Q9ALG9	Q9alg9 uncultured		5	10.9
892	120	2	Q9ALG8	Q9alg8 uncultured		5	10.9
893	120	2	Q9ALG7	Q9alg7 uncultured		5	10.9
894	120	2	Q9ALG4	Q9alg4 uncultured		5	10.9
895	120	2	Q9ALG1	Q9alg1 uncultured		5	10.9
896	120	2	Q9ALG0	Q9algo uncultured		5	10.9
897	120	2	Q9ALF9	Q9alf9 uncultured		5	10.9
898	120	2	Q9ALF7	Q9alf7 uncultured		5	10.9
899	120	2	Q9ALF5	Q9alf5 uncultured		5	10.9
900	120	2	Q9ALE9	Q9ale9 uncultured		5	10.9
901	120	2	Q9ALE8	Q9ale8 uncultured		5	10.9
902	120	2	Q9ALE7	Q9ale7 uncultured		5	10.9
903	120	2	Q9ALE5	Q9ale5 uncultured		5	10.9
904	120	2	Q9ALE3	Q9ale3 uncultured		5	10.9
905	120	2	Q9ALE0	Q9ale0 uncultured		5	10.9
906	120	2	Q9ALD9	Q9ald9 uncultured		5	10.9
907	120	2	Q9ALD8	Q9ald8 uncultured		5	10.9
908	120	2	Q9ALD7	Q9ald7 uncultured		5	10.9
909	120	2	Q9ALD6	Q9ald6 uncultured		5	10.9
910	120	2	Q9ALD5	Q9ald5 uncultured		5	10.9
911	120	2	Q9ALD4	Q9ald4 uncultured		5	10.9
912	120	2	Q9ALD3	Q9ald3 uncultured		5	10.9
913	120	2	Q9ALD2	Q9ald2 uncultured		5	10.9
914	120	2	Q9ALD1	Q9ald1 uncultured		5	10.9
915	120	2	Q9ALC6	Q9alc6 uncultured		5	10.9
916	120	2	Q9ALC1	Q9alc1 uncultured		5	10.9
917	120	2	Q9ALC0	Q9alc0 uncultured		5	10.9
918	120	2	Q9ALB9	Q9alb9 uncultured		5	10.9
919	120	2	Q9ALB8	Q9alb8 uncultured		5	10.9
920	120	2	Q9ALB7	Q9alb7 uncultured		5	10.9
921	120	2	Q9ALB1	Q9alb1 uncultured		5	10.9
922	120	2	Q93TH2	Q93th2 uncultured		5	10.9
923	120	2	Q93L61	Q93l61 bacteroides		5	10.9
924	120	2	Q93KN6	Q93kn6 uncultured		5	10.9
925	120	2	Q9NAS2	Q9nas2 brugia mala		5	10.9
926	120	17	O58311	O58311 pyrococcus		5	10.9
927	120	17	O58601	O58601 pyrococcus		5	10.9
928	121	2	Q9RNU0	Q9rnu0 solar lake		5	10.9
929	121	2	Q9RNT6	Q9rnt6 solar lake		5	10.9
930	121	2	Q9RNT4	Q9rnt4 solar lake		5	10.9
931	121	2	Q9RNT3	Q9rnt3 solar lake		5	10.9
932	121	2	Q9RNT1	Q9rnt1 solar lake		5	10.9
933	121	2	Q9RNS9	Q9rns9 solar lake		5	10.9
934	121	2	Q9RNS7	Q9rns7 solar lake		5	10.9
935	121	2	Q9RNS5	Q9rns5 solar lake		5	10.9
936	121	2	Q9RNS4	Q9rns4 solar lake		5	10.9
937	121	2	Q9RNS2	Q9rns2 solar lake		5	10.9
938	121	2	Q9RNS1	Q9rns1 solar lake		5	10.9
939	121	2	Q9RMR3	Q9rmr3 solar lake		5	10.9
940	121	2	Q9RMR1	Q9rmr1 solar lake		5	10.9
941	121	2	Q9ALH8	Q9alh8 uncultured		5	10.9
942	121	2	Q9ALH5	Q9alh5 uncultured		5	10.9
943	121	2	Q9ALG6	Q9alg6 uncultured		5	10.9
944	121	2	Q9ALF8	Q9alf8 uncultured		5	10.9
945	121	2	Q9ALF3	Q9alf3 uncultured		5	10.9
946	121	2	Q9ALF1	Q9alf1 uncultured		5	10.9
947	121	2	Q9ALF0	Q9alf0 uncultured		5	10.9
948	121	2	Q9ALD0	Q9ald0 uncultured		5	10.9
949	121	2	Q9ALC5	Q9alc5 uncultured		5	10.9
950	121	2	Q9ALC4	Q9alc4 uncultured		5	10.9
951	121	2	Q9ALC2	Q9alc2 uncultured		5	10.9
952	121	2	Q9ALB5	Q9alb5 uncultured		5	10.9
953	121	2	Q9ALB4	Q9alb4 uncultured		5	10.9
954	121	2	Q9ALB3	Q9alb3 uncultured		5	10.9
955	121	2	Q9ALB2	Q9alb2 uncultured		5	10.9
956	121	5	P91789	P91789 taenia ovis		5	10.9
957	121	5	Q95YV8	Q95yv8 leishmania		5	10.9
958	121	12	O41658	O41658 influenza a		5	10.9
959	121	15	Q9WNY0	Q9wny0 human immun		5	10.9
960	122	2	Q9RNT7	Q9rnt7 solar lake		5	10.9
961	122	5	Q9RNS3	Q9rns3 solar lake		5	10.9
962	122	5	Q9TYG7	Q9tyg7 plectus acu		5	10.9
963	122	16	Q9JTV4	Q9jtv4 neisseria m		5	10.9
964	123	5	Q9ALB6	Q9alb6 uncultured		5	10.9
965	123	5	Q9V8Q4	Q9v8q4 drosophila		5	10.9

	Accession				Accession		
966	Q99QI5	5	10.9	Q40797	physalis cr	123	10
967	Q99QI5	5	10.9	Q914N7	pseudomonas	123	16
968	Q99QI5	5	10.9	Q9yak9	aeropyrum p	123	17
969	Q99QI5	5	10.9	Q9ya10	aeropyrum p	123	17
970	Q99QI5	5	10.9	Q9xy31	drosophila	124	5
971	Q99QI5	5	10.9	Q9xy30	drosophila	124	5
972	Q99QI5	5	10.9	Q9xy29	drosophila	124	5
973	Q99QI5	5	10.9	Q9xy28	drosophila	124	5
974	Q99QI5	5	10.9	Q9xy27	drosophila	124	5
975	Q99QI5	5	10.9	Q9xy26	drosophila	124	5
976	Q99QI5	5	10.9	Q9xy25	drosophila	124	5
977	Q99QI5	5	10.9	Q9xy24	drosophila	124	5
978	Q99QI5	5	10.9	Q9xy23	drosophila	124	5
979	Q99QI5	5	10.9	Q9xy22	drosophila	124	5
980	Q99QI5	5	10.9	Q9vm01	drosophila	124	5
981	Q99QI5	5	10.9	Q9niv4	norrisia no	124	5
982	Q99QI5	5	10.9	Q9qzv0	rattus norv	124	11
983	Q99QI5	5	10.9	Q9dfk3	gillichthys	124	13
984	Q99QI5	5	10.9	O14037	schizosacch	125	3
985	Q99QI5	5	10.9	Q9gqy9	plasmodium	125	5
986	Q99QI5	5	10.9	Q9ctk3	mus musculu	125	11
987	Q99QI5	5	10.9	Q60911	mus musculu	125	11
988	Q99QI5	5	10.9	Q912y5	mus musculu	125	11
989	Q99QI5	5	10.9	Q9dhr0	yaba-like d	125	12
990	Q99QI5	5	10.9	Q9ddl4	callisaurus	125	13
991	Q99QI5	5	10.9	Q9hww9	pseudomonas	125	16
992	Q99QI5	5	10.9	Q92b00	listeria in	125	16
993	Q99QI5	5	10.9	Q9h7a1	homo sapien	126	4
994	Q99QI5	5	10.9	O39627	cydia pomon	126	12
995	Q99QI5	5	10.9	O89457	heliothis a	126	12
996	Q99QI5	5	10.9	Q91mv4	lumpy skin	126	12
997	Q99QI5	5	10.9	Q9vd37	drosophila	127	5
998	Q99QI5	5	10.9	Q9nlm1	apis mellif	127	5
999	Q99QI5	5	10.9	Q9g8z6	ochromonas	127	8
1000	Q99QI5	5	10.9	P70370	mus musculu	127	11

ALIGNMENTS

RESULT 1
Q99QI5
ID   Q99QI5   PRELIMINARY;   PRT;   46 AA.
AC   Q99QI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN PRECURSOR.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB14, H7, LT11, NG8, AND UA159;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277152; AAK01542.1; -.
DR   EMBL; AF277153; AAK01543.1; -.
DR   EMBL; AF277155; AAK01545.1; -.
DR   EMBL; AF277156; AAK01546.1; -.
DR   EMBL; AF277157; AAK01547.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
FT   CHAIN   26   46   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA; 5211 MW; 38FA62B6F78FC3BF CRC64;

Query Match   100.0%; Score 46; DB 2; Length 46;
Best Local Similarity   100.0%; Pred. No. 3.5e-40;
Matches   46; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46

RESULT 2
Q9APK7
ID   Q9APK7   PRELIMINARY;   PRT;   46 AA.
AC   Q9APK7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BM71;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277151; AAK01541.1; -.
FT   CHAIN   26   46   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA; 5195 MW; 3E0B9DB5B8FC3BF CRC64;

Query Match   89.1%; Score 41; DB 2; Length 46;
Best Local Similarity   100.0%; Pred. No. 5.1e-35;
Matches   41; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   6 SLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
         |||||||||||||||||||||||||||||||||||||||||
Db   6 SLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46

RESULT 3
Q9APK6
ID   Q9APK6   PRELIMINARY;   PRT;   43 AA.
AC   Q9APK6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH1005;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277154; AAK01544.1; -.
FT   CHAIN   26   43   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   43 AA; 4927 MW; E6A78FC3BF6156C7 CRC64;

Query Match   58.7%; Score 27; DB 2; Length 43;
Best Local Similarity   100.0%; Pred. No. 1.3e-20;
Matches   27; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKKTLSLKNDFKEIKTDELEIIIGGSG 27
         |||||||||||||||||||||||||||
Db   1 MKKTLSLKNDFKEIKTDELEIIIGGSG 27

```
RESULT 4
Q9R7H6
ID   Q9R7H6          PRELIMINARY;      PRT;    363 AA.
AC   Q9R7H6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 41.9 KDA PROTEIN (FRAGMENT).
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N40;
RA   Feng S., Chen G., Barthold S.;
RT   "G4B clone from Borrelia burgdorferi.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF005055; AAD01225.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   363 AA;  41909 MW;  D833015494093ZCA CRC64;

Query Match             19.6%;  Score 9;  DB 2;  Length 363;
Best Local Similarity   100.0%;  Pred. No. 0.31;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 TLSLKNDFK 12
          |||||||||
Db     45 TLSLKNDFK 53

RESULT 5
Q98RG2
ID   Q98RG2          PRELIMINARY;      PRT;     88 AA.
AC   Q98RG2;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL PROTEIN MYPU_0470.
GN   MYPU_0470.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445563; CAC13220.1; -.
DR   MypuList; MYPU_0470; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   88 AA;  10502 MW;  49ECE9BD060A3407 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Length 88;
Best Local Similarity   100.0%;  Pred. No. 10;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 KTLSLKN 9
          |||||||
Db     20 KTLSLKN 26

RESULT 6
Q83163
ID   Q83163          PRELIMINARY;      PRT;     96 AA.
AC   Q83163;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   (ALTERED VIRULENCE ISOLATE D/H), COMPLETE GENOME.
OS   Cauliflower mosaic virus.
OC   Viruses; Retroid viruses; Caulimovirus.
OX   NCBI_TaxID=10641;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CABB-D/H;
RX   MEDLINE=83106468; PubMed=7152260;
RA   Balazs E., Guilly H., Jonard G., Richards K.;
RT   "Nucleotide sequence of DNA from an altered-virulence isolate D/H of
RT   the cauliflower mosaic virus.";
RL   Gene 19:239-249(1982).
DR   EMBL; M10376; AAA6344.1; -.
SQ   SEQUENCE   96 AA;  11429 MW;  F43565C5790A7657 CRC64;

Query Match             15.2%;  Score 7;  DB 12;  Length 96;
Best Local Similarity   100.0%;  Pred. No. 11;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 KTLSLKN 9
          |||||||
Db     75 KTLSLKN 81

RESULT 7
Q96Y40
ID   Q96Y40          PRELIMINARY;      PRT;    239 AA.
AC   Q96Y40;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN ST2326.
GN   ST2326.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
DR   EMBL; AP000989; BAB67437.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   239 AA;  28001 MW;  1718022FE4D067EC CRC64;

Query Match             15.2%;  Score 7;  DB 17;  Length 239;
Best Local Similarity   100.0%;  Pred. No. 25;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 SLKNDFK 12
          |||||||
Db    148 SLKNDFK 154

RESULT 8
Q9I488
ID   Q9I488          PRELIMINARY;      PRT;    240 AA.
AC   Q9I488;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
```

```
DE   PROBABLE ATP-BINDING COMPONENT OF ABC TRANSPORTER.
GN   PA1256.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS).
DR   EMBL; AE004555; AAG04645.1; -.
DR   InterPro; IPR003593; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Complete proteome; Transport.
SQ   SEQUENCE   240 AA;   26548 MW;   3F6796D7C10F0F83 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Length 240;
  Best Local Similarity   100.0%;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         22 IIGGSGS 28
              |||||||
Db         32 IIGGSGS 38
```

```
RESULT 9
Q92WS3
ID   Q92WS3;    PRELIMINARY;    PRT;   240 AA.
AC   Q92WS3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE AMINO ACID ABC TRANSPORTER ATP-BINDING PROTEIN.
GN   SMB20266.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymB (megaplasmid 2).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396508; PubMed=11481431;
RA   Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA   Vorhoelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA   Golding B., Puehler A.;
RT   "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT   fixing endosymbiont Sinorhizobium meliloti.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR   EMBL; AL603642; CAC48656.1; -.
KW   ATP-binding; Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   240 AA;   26366 MW;   E9826924BC60E597 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Length 240;
  Best Local Similarity   100.0%;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         22 IIGGSGS 28
              |||||||
Db         32 IIGGSGS 38
```

```
RESULT 10
Q90315
ID   Q90315;    PRELIMINARY;    PRT;   253 AA.
AC   Q90315;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STEM CELL FACTOR, MEMBRANE-BOUND FORM.
OS   Coturnix coturnix (common quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=9091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96283808; PubMed=8679698;
RA   Petitte J.N., Kulik M.J.;
RT   "Cloning and characterization of cDNAs encoding two forms of avian
RT   stem cell factor.";
RL   Biochim. Biophys. Acta 1307:149-151(1996).
DR   EMBL; U43079; AAC59934.1; -.
DR   InterPro; IPR003452; SCF.
DR   Pfam; PF02404; SCF; 1.
SQ   SEQUENCE   253 AA;   28993 MW;   20709B3854F0207A CRC64;

Query Match             15.2%;  Score 7;  DB 13;  Length 253;
  Best Local Similarity   100.0%;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 FFRLFNR 38
              |||||||
Db        144 FFRLFNR 150
```

```
RESULT 11
O67894
ID   O67894;    PRELIMINARY;    PRT;   253 AA.
AC   O67894;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ABC TRANSPORTER.
GN   ABCT12 OR AQ_2137.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000774; AAC07853.1; -.
HSSP; Q58663; 1G6H.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
KW   Complete proteome.
SQ   SEQUENCE   253 AA;   28621 MW;   C880AE769CF8A4B2 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Length 253;
  Best Local Similarity   100.0%;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

Wed Nov 13 14:19:05 2002   us-09-833-017b-2.oligo.rspt   Page 11

```
Qy      22 IIGGSGS 28
           |||||||
Db      35 IIGGSGS 41

RESULT 12
Q98CC9
ID   Q98CC9;                      PRELIMINARY;       PRT;     268 AA.
AC   Q98CC9;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   AMINO ACID ABC TRANSPORTER, ATP-BINDING PROTEIN.
GN   MLL5204.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa K., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003006; BAB51692.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
KW   ATP-binding; Complete proteome.
SQ   SEQUENCE   268 AA;  29207 MW;  8676DDE8ED78FD34 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Pred. No. 28;  Length 268;
Best Local Similarity   100.0%;
Matches     7;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy      22 IIGGSGS 28
           |||||||
Db      49 IIGGSGS 55

RESULT 13
Q9SEW3
ID   Q9SEW3;                      PRELIMINARY;       PRT;     276 AA.
AC   Q9SEW3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RECEPTOR-LIKE PROTEIN KINASE (FRAGMENT).
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. WILLIAMS 82;
RA   Bassuner R., Cao H., Mackenzie S.;
RT   "A novel receptor-like protein kinase from soybean and its relation to
RT   the resistance phenotype to cyst nematode infection.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF068135; AAF21775.1; -.
DR   HSSP; P05046; 1SBE.
DR   InterPro; IPR000985; Lectin_legA.
DR   Pfam; PF00138; lectin_legA; 1.
```

```
KW   Kinase.
FT   NON_TER       1      1
FT   NON_TER     276    276
SQ   SEQUENCE   276 AA;  30521 MW;  FAFEB0769E0D1114 CRC64;

Query Match             15.2%;  Score 7;  DB 10;  Length 276;
Best Local Similarity   100.0%;              Pred. No. 28;
Matches     7;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       3 KTLSLKN 9
           |||||||
Db     226 KTLSLKN 232

RESULT 14
Q92S62
ID   Q92S62;                      PRELIMINARY;       PRT;     284 AA.
AC   Q92S62;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE ATP-BINDING ABC TRANSPORTER PROTEIN.
GN   SMC02260.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21368234; PubMed=11474104;
RA   Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA   Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA   Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA   Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA   Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA   Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA   Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA   Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA   Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA   Vorhoelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT   "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL   Science 293:668-672(2001).
DR   EMBL; AL591784; CAC45136.1; -.
KW   Complete proteome.
SQ   SEQUENCE   284 AA;  31218 MW;  82E1AFB4DAA493C1 CRC64;

Query Match             15.2%;  Score 7;  DB 16;  Length 284;
Best Local Similarity   100.0%;              Pred. No. 29;
Matches     7;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy      22 IIGGSGS 28
           |||||||
Db      65 IIGGSGS 71

RESULT 15
Q90314
ID   Q90314;                      PRELIMINARY;       PRT;     287 AA.
AC   Q90314;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STEM CELL FACTOR.
OS   Coturnix coturnix (common quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=9091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96283808; PubMed=8679698;
RA   Petitte J.N., Kulik M.J.;
```

RT   "Cloning and characterization of cDNAs encoding two forms of avian
RT   stem cell factor.";
RL   Biochim. Biophys. Acta 1307:149-151(1996).
DR   EMBL; U43078; AAC59933.1; -.
DR   InterPro; IPR003452; SCF.
DR   Pfam; PF02404; SCF; 1.
SQ   SEQUENCE   287 AA; 32455 MW; ABA81AEA422A702E CRC64;

Query Match   15.2%; Score 7; DB 13; Length 287;
Best Local Similarity 100.0%; Pred. No. 29;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   32 FFRLFNR 38
          |||||||
Db   144 FFRLFNR 150

Search completed: November 5, 2002, 11:06:28
Job time : 87.209 secs

```
                    GenCore version 5.1.3
          Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    November  5, 2002, 10:55:02 ; Search time 16.9254 Seconds
                                          (without alignments)
                               137.814 Million cell updates/sec

Title:          US-09-833-017B-4
Perfect score:  104
Sequence:       1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters:    747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      A_Geneseq_032802:*
          1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
          2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
          3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
          4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
          5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
          6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
          7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
          8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
          9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
         10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
         11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
         12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
         13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
         14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
         15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
         16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
         17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
         18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
         19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
         20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
         21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
         22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	44.2	577	21	AAG48997	Arabidopsis thalia
2	46	44.2	596	21	AAG48996	Arabidopsis thalia
3	46	44.2	660	21	AAG48995	Arabidopsis thalia
4	45	43.3	391	22	AAB70114	Murine 3-OST-3B.
5	44	42.3	340	22	AAF09656	Human gene 18 enco
6	44	42.3	340	22	AAU18677	Renal and cardiova
7	44	42.3	340	22	AAU18712	Renal and cardiova
8	44	42.3	719	22	AAB95255	Human protein sequ
9	44	41.3	86	22	AAO07963	Human polypeptide
10	43	41.3	106	19	AAW79359	DNA encoding Staph
11	43	41.3	135	14	AAR41945	Serotonin receptor
12	43	41.3	151	19	AAW79358	DNA encoding Staph
13	43	41.3	167	22	AAU34248	Staphylococcus aur
14	43	41.3	167	22	AAU36819	Staphylococcus aur
15	43	41.3	167	22	AAU37429	Staphylococcus aur
16	43	41.3	479	14	AAR41942	Serotonin receptor
17	43	41.3	504	15	AAR54682	Mouse brain 5HT2C
18	42	40.4	43	22	AAO06271	Human polypeptide
19	42	40.4	114	22	AAB95669	Human protein sequ
20	42	40.4	148	22	AAB93004	Human protein sequ
21	42	40.4	225	22	ABB65926	Drosophila melanog
22	42	40.4	443	21	ABB42806	Human ORFX ORF2570
23	42	40.4	453	21	AAY68799	Amino acid sequenc
24	42	40.4	482	22	AAU31776	Novel human secret
25	42	40.4	499	22	ABB62187	Drosophila melanog
26	41	39.4	148	22	ABG28821	Novel human diagno
27	41	39.4	183	18	AAW28302	Amino acid sequenc
28	41	39.4	250	21	AAB58800	Breast and ovarian
29	41	39.4	437	19	AAW53264	Human adipocyte-sp
30	41	39.4	437	21	AAY99534	Human adipocyte-sp
31	41	39.4	595	20	AAY27236	Amino acid sequenc
32	41	39.4	619	20	AAY27235	Amino acid sequenc
33	41	39.4	761	20	AAY27211	Amino acid sequenc
34	41	39.4	1798	22	ABG09933	Novel human diagno
35	41	39.4	1981	19	AAW42634	Protein sequence
36	41	39.4	2001	22	AAY72649	Human ATP binding
37	41	39.4	2436	22	AAG67160	Amino acid sequenc
38	41	39.4	2436	22	AAB62210	Human ABCA2 transp
39	40	38.5	48	22	AAG76591	Human colon cancer
40	40	38.5	52	21	AAB43141	Human ORFX ORF2905
41	40	38.5	70	22	AAU02768	Mutant SCF polypep
42	40	38.5	70	22	AAB73569	Mutant SCF polypep
43	40	38.5	82	22	AAU05264	Mutant SCF polypep
44	40	38.5	82	22	AAB98066	Human SCF related
45	40	38.5	82	22	AAU02459	Mutant SCF polypep

ALIGNMENTS

```
RESULT 1
AAG48997
ID   AAG48997 standard; Protein; 577 AA.
XX
AC   AAG48997;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID, NO: 61937.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
```

```
PR  23-APR-1999;  99US-0130510.
PR  23-APR-1999;  99US-0130891.
PR  28-APR-1999;  99US-0131449.
PR  30-APR-1999;  99US-0132048.
PR  30-APR-1999;  99US-0132407.
PR  04-MAY-1999;  99US-0132484.
PR  05-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  06-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  18-MAY-1999;  99US-0134370.
PR  19-MAY-1999;  99US-0134768.
PR  20-MAY-1999;  99US-0134941.
PR  21-MAY-1999;  99US-0135124.
PR  24-MAY-1999;  99US-0135353.
PR  25-MAY-1999;  99US-0135629.
PR  27-MAY-1999;  99US-0136021.
PR  28-MAY-1999;  99US-0136392.
PR  01-JUN-1999;  99US-0136782.
PR  03-JUN-1999;  99US-0137222.
PR  04-JUN-1999;  99US-0137528.
PR  07-JUN-1999;  99US-0137502.
PR  08-JUN-1999;  99US-0137724.
PR  10-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  14-JUN-1999;  99US-0138847.
PR  16-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  17-JUN-1999;  99US-0139453.
PR  18-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  21-JUN-1999;  99US-0139763.
PR  22-JUN-1999;  99US-0139817.
PR  23-JUN-1999;  99US-0139899.
PR  24-JUN-1999;  99US-0140353.
PR  28-JUN-1999;  99US-0140354.
PR  29-JUN-1999;  99US-0140695.
PR  30-JUN-1999;  99US-0140823.
PR  01-JUL-1999;  99US-0140991.
PR  01-JUL-1999;  99US-0141287.
PR  02-JUL-1999;  99US-0141842.
PR  06-JUL-1999;  99US-0142154.
PR  08-JUL-1999;  99US-0142055.
PR  09-JUL-1999;  99US-0142390.
PR  12-JUL-1999;  99US-0142803.
PR  13-JUL-1999;  99US-0142920.
PR  14-JUL-1999;  99US-0142977.
PR  15-JUL-1999;  99US-0143542.
PR  16-JUL-1999;  99US-0143624.
PR  16-JUL-1999;  99US-0144005.
PR  19-JUL-1999;  99US-0144085.
PR  19-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  20-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.

PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  23-JUL-1999;  99US-0145089.
PR  23-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  26-JUL-1999;  99US-0145218.
PR  27-JUL-1999;  99US-0145224.
PR  27-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  28-JUL-1999;  99US-0145918.
PR  02-AUG-1999;  99US-0145919.
PR  02-AUG-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  03-AUG-1999;  99US-0146388.
PR  04-AUG-1999;  99US-0146389.
PR  04-AUG-1999;  99US-0147038.
PR  05-AUG-1999;  99US-0147204.
PR  05-AUG-1999;  99US-0147302.
PR  06-AUG-1999;  99US-0147192.
PR  06-AUG-1999;  99US-0147260.
PR  09-AUG-1999;  99US-0147303.
PR  09-AUG-1999;  99US-0147416.
PR  10-AUG-1999;  99US-0147493.
PR  11-AUG-1999;  99US-0147935.
PR  12-AUG-1999;  99US-0148171.
PR  13-AUG-1999;  99US-0148319.
PR  13-AUG-1999;  99US-0148341.
PR  16-AUG-1999;  99US-0148565.
PR  17-AUG-1999;  99US-0148684.
PR  18-AUG-1999;  99US-0149368.
PR  20-AUG-1999;  99US-0149175.
PR  20-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  23-AUG-1999;  99US-0149723.
PR  23-AUG-1999;  99US-0149929.
PR  25-AUG-1999;  99US-0149902.
PR  26-AUG-1999;  99US-0149930.
PR  27-AUG-1999;  99US-0150566.
PR  27-AUG-1999;  99US-0150084.
PR  27-AUG-1999;  99US-0151065.
PR  30-AUG-1999;  99US-0151066.
PR  31-AUG-1999;  99US-0251080.
PR  01-SEP-1999;  99US-0151303.
PR  07-SEP-1999;  99US-0151438.
PR  10-SEP-1999;  99US-0151930.
PR  13-SEP-1999;  99US-0153070.
PR  15-SEP-1999;  99US-0153363.
PR  16-SEP-1999;  99US-0153758.
PR  20-SEP-1999;  99US-0154018.
PR  22-SEP-1999;  99US-0154039.
PR  23-SEP-1999;  99US-0154779.
PR  24-SEP-1999;  99US-0155139.
PR  28-SEP-1999;  99US-0155486.
PR  29-SEP-1999;  99US-0155659.
PR  04-OCT-1999;  99US-0156458.
PR  05-OCT-1999;  99US-0156596.
PR  06-OCT-1999;  99US-0157117.
PR  07-OCT-1999;  99US-0157753.
PR  08-OCT-1999;  99US-0157865.
PR  12-OCT-1999;  99US-0158029.
PR  13-OCT-1999;  99US-0158232.
PR  13-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  14-OCT-1999;  99US-0159294.
PR  14-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
```

```
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.


Query Match             44.2%;  Score 46; DB 21; Length 577;
Best Local Similarity   69.2%;  Pred. No. 24;
Matches    9;  Conservative   1;  Mismatches   3;  Indels   0;  Gaps    0;

Qy      7 FFRLFNRSFTQAL 19
          |||||:||:||| |
Db    411 FFRLFDRDFRQRL 423


RESULT 2
AAG48996
ID   AAG48996 standard; Protein; 596 AA.
XX
AC   AAG48996;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 61936.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0128845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   29-MAR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   07-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145087.
```

```
PR   22-JUL-1999;  99US-0145089.          PR   21-OCT-1999;  99US-0160814.
PR   22-JUL-1999;  99US-0145192.          PR   21-OCT-1999;  99US-0160815.
PR   23-JUL-1999;  99US-0145145.          PR   22-OCT-1999;  99US-0160980.
PR   23-JUL-1999;  99US-0145218.          PR   22-OCT-1999;  99US-0160981.
PR   23-JUL-1999;  99US-0145224.          PR   22-OCT-1999;  99US-0160989.
PR   26-JUL-1999;  99US-0145276.          PR   25-OCT-1999;  99US-0161404.
PR   27-JUL-1999;  99US-0145913.          PR   25-OCT-1999;  99US-0161405.
PR   27-JUL-1999;  99US-0145918.          PR   25-OCT-1999;  99US-0161406.
PR   27-JUL-1999;  99US-0145919.          PR   26-OCT-1999;  99US-0161359.
PR   28-JUL-1999;  99US-0145951.          PR   26-OCT-1999;  99US-0161360.
PR   02-AUG-1999;  99US-0146386.          PR   26-OCT-1999;  99US-0161361.
PR   02-AUG-1999;  99US-0146388.          PR   28-OCT-1999;  99US-0161920.
PR   02-AUG-1999;  99US-0146389.          PR   28-OCT-1999;  99US-0161992.
PR   03-AUG-1999;  99US-0147038.          PR   28-OCT-1999;  99US-0161993.
PR   04-AUG-1999;  99US-0147204.          PR   29-OCT-1999;  99US-0162142.
PR   04-AUG-1999;  99US-0147302.
PR   05-AUG-1999;  99US-0147192.          Query Match          44.2%;    Score 46;  DB 21;  Length 596;
PR   05-AUG-1999;  99US-0147260.          Best Local Similarity 69.2%;    Pred. No. 25;
PR   06-AUG-1999;  99US-0147303.          Matches    9;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;
PR   06-AUG-1999;  99US-0147416.
PR   09-AUG-1999;  99US-0147493.          Qy      7  FFRLFNRSFTQAL  19
PR   09-AUG-1999;  99US-0147935.                     |||||:||||:|||
PR   10-AUG-1999;  99US-0148171.          Db    430  FFRLFDRDFRQRL  442
PR   11-AUG-1999;  99US-0148319.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.          RESULT 3
PR   16-AUG-1999;  99US-0149368.          AAG48995
PR   17-AUG-1999;  99US-0149175.          ID   AAG48995 standard; Protein; 660 AA.
PR   18-AUG-1999;  99US-0149426.          XX
PR   20-AUG-1999;  99US-0149722.          AC   AAG48995;
PR   20-AUG-1999;  99US-0149723.          XX
PR   20-AUG-1999;  99US-0149929.          DT   18-OCT-2000  (first entry)
PR   23-AUG-1999;  99US-0149902.          XX
PR   23-AUG-1999;  99US-0149930.          DE   Arabidopsis thaliana protein fragment SEQ ID NO: 61935.
PR   25-AUG-1999;  99US-0150566.          XX
PR   26-AUG-1999;  99US-0150884.          KW   Protein identification; signal transduction pathway; metabolic pathway;
PR   27-AUG-1999;  99US-0151065.          KW   hybridisation assay; genetic mapping; gene expression control; promoter;
PR   27-AUG-1999;  99US-0151066.          KW   termination sequence.
PR   27-AUG-1999;  99US-0151080.          XX
PR   30-AUG-1999;  99US-0151303.          OS   Arabidopsis thaliana.
PR   31-AUG-1999;  99US-0151438.          XX
PR   01-SEP-1999;  99US-0151930.          PN   EP1033405-A2.
PR   07-SEP-1999;  99US-0152363.          XX
PR   10-SEP-1999;  99US-0153070.          PD   06-SEP-2000.
PR   13-SEP-1999;  99US-0153758.          XX
PR   15-SEP-1999;  99US-0154018.          PF   25-FEB-2000;  2000EP-0301439.
PR   16-SEP-1999;  99US-0154039.          XX
PR   20-SEP-1999;  99US-0154779.          PR   25-FEB-1999;  99US-0121825.
PR   22-SEP-1999;  99US-0155139.          PR   05-MAR-1999;  99US-0123180.
PR   23-SEP-1999;  99US-0155486.          PR   09-MAR-1999;  99US-0123548.
PR   24-SEP-1999;  99US-0155659.          PR   23-MAR-1999;  99US-0125788.
PR   28-SEP-1999;  99US-0156458.          PR   25-MAR-1999;  99US-0126264.
PR   29-SEP-1999;  99US-0156596.          PR   29-MAR-1999;  99US-0267785.
PR   04-OCT-1999;  99US-0157117.          PR   01-APR-1999;  99US-0127462.
PR   05-OCT-1999;  99US-0157753.          PR   06-APR-1999;  99US-0128234.
PR   06-OCT-1999;  99US-0157865.          PR   08-APR-1999;  99US-0128714.
PR   07-OCT-1999;  99US-0158029.          PR   16-APR-1999;  99US-0129845.
PR   08-OCT-1999;  99US-0158232.          PR   19-APR-1999;  99US-0130077.
PR   12-OCT-1999;  99US-0158369.          PR   21-APR-1999;  99US-0130449.
PR   13-OCT-1999;  99US-0159293.          PR   23-APR-1999;  99US-0130510.
PR   13-OCT-1999;  99US-0159294.          PR   23-APR-1999;  99US-0130891.
PR   13-OCT-1999;  99US-0159295.          PR   28-APR-1999;  99US-0131449.
PR   14-OCT-1999;  99US-0159329.          PR   30-APR-1999;  99US-0132048.
PR   14-OCT-1999;  99US-0159330.          PR   30-APR-1999;  99US-0132407.
PR   14-OCT-1999;  99US-0159331.          PR   04-MAY-1999;  99US-0132484.
PR   14-OCT-1999;  99US-0159637.          PR   05-MAY-1999;  99US-0132485.
PR   14-OCT-1999;  99US-0159638.          PR   06-MAY-1999;  99US-0132486.
PR   18-OCT-1999;  99US-0159584.          PR   06-MAY-1999;  99US-0132487.
PR   21-OCT-1999;  99US-0160741.          PR   07-MAY-1999;  99US-0132863.
PR   21-OCT-1999;  99US-0160767.          PR   11-MAY-1999;  99US-0134256.
PR   21-OCT-1999;  99US-0160768.          PR   14-MAY-1999;  99US-0134218.
PR   21-OCT-1999;  99US-0160770.          PR   14-MAY-1999;  99US-0134219.
                                          PR   14-MAY-1999;  99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   44.2%; Score 46; DB 21; Length 660;
Best Local Similarity 69.2%; Pred. No. 28;
Matches   9; Conservative   1; Mismatches   3; Indels   0; Gaps   0;

Qy   7 FFRLFNRSFTQAL 19
           ||||| :|| | |
Db   494 FFRLFDRDFRQRL 506

RESULT 4
AAB70114
ID   AAB70114 standard; protein; 391 AA.
XX
AC   AAB70114;
XX
DT   17-MAY-2001 (first entry)
XX
DE   Murine 3-OST-3B.
XX
KW   Mouse; 3-OST-3B; heparan sulphate 3-O-sulphotransferase 3B; antiviral;
KW   herpes simplex virus type-1; HSV-1; infection.
XX
OS   Mus sp.
XX
PN   WO200113910-A2.
XX
PD   01-MAR-2001.
XX
PF   25-AUG-2000; 2000WO-US23288.
XX
PR   25-AUG-1999; 99US-0150743.
PR   10-JAN-2000; 2000US-0175347.
XX
PA   (MASI) MASSACHUSETTS INST TECHNOLOGY.
PA   (NOUN) UNIV NORTHWESTERN.
XX
PI   Shukla D, Liu J, Rosenberg RD, Spear PG;
XX
DR   WPI; 2001-234971/24.
XX
PT   Novel pharmaceutical preparation useful for inhibiting herpes simplex
PT   virus type-1 infection in a mammal, comprises a polysaccharide
PT   preparation enriched for 3-O-sulfated glycosamine residues -
XX
PS   Disclosure: Fig 1; 29pp; English.
XX
CC   The present sequence is murine heparan sulphate 3-O-sulphotransferase
CC   (3-OST)-3B. A substantially pure polysaccharide preparation enriched for
CC   3-OST-3 modified glucosamine residues is useful for inhibiting herpes
CC   simplex virus type-1 (HSV-1) infection in a mammal. It is useful for
CC   inhibiting viral penetration and entry of HSV-1 into mammalian cells
CC   through heparan sulphate binding.
XX
SQ   Sequence   391 AA;

Query Match   43.3%; Score 45; DB 22; Length 391;
Best Local Similarity 50.0%; Pred. No. 23;
Matches   9; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   4 LSTFFRLFNRSFTQALGK 21
           | |:|||||||| |:
Db   369 LRDFYRPFNRKFYQMTGR 386

RESULT 5
AAE09656
ID   AAE09656 standard; protein; 340 AA.
XX
AC   AAE09656;
XX
DT   19-NOV-2001 (first entry)
XX
DE   Human gene 18 encoded lipid metabolism protein HDACA29, SEQ ID NO:50.
XX
KW   Human; lipid metabolism protein; nootropic; neuroprotective; cardiant;
KW   cerebroprotective; hepatotropic; antidiabetic; ophthalmic; nephrotropic;
KW   immune disorder; autoimmune disease; rheumatoid arthritis; glossitis;
KW   systemic lupus erythematosus; conjunctivitis; inflammatory disorder;
KW   respiratory disorder; asthma; allergy; CNS disorder; Alzheimer's disease;
KW   Parkinson's disease; atherosclerosis; cardiovascular disorder; cancer;
KW   coronary disease; familial hypercholesterolaemia; hyperlipidaemia;
KW   haematopoietic disorder; hypolipidaemia; lipidosis; Gaucher's disease;
KW   Tay-sach's disease; mental retardation; gene therapy; antisense therapy.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 161
FT   /label= Unknown
FT   /note= "Encoded by MAA"
FT   Misc-difference 278
FT   /label= Unknown
FT   /note= "Encoded by AAN"
FT   Misc-difference 293
FT   /label= Unknown
FT   /note= "Encoded by GAW"
XX
PN   WO200155203-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01327.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.

```
PR   23-AUG-2000;   2000US-0227009.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0233398.
PR   14-SEP-2000;   2000US-0233399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   26-SEP-2000;   2000US-0234998.
PR   27-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0238802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.

PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   06-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-45716/49.
DR   N-PSDB; AAD16623.
XX
PT   Isolated lipid metabolism polypeptide for screening to identify
PT   antagonists and agonists that may enhance or block activities mediated
PT   by lipid metabolism proteins and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Claim 11; SEQ ID No 50; 453pp; English.
XX
CC   AAD16606-AAD16627 represent cDNAs corresponding to human lipid metabolism
CC   protein genes, and AAE09639-AAE09660 represent the proteins they encode.
CC   AAD16628-AAD16647 represent human lipid metabolism genomic DNAs. The
CC   genes and their corresponding proteins are useful for preventing,
CC   treating or ameliorating medical conditions, such as immune disorders,
CC   inflammatory disorders (e.g., conjunctivitis, glossitis), autoimmune
CC   diseases (e.g., rheumatoid arthritis, systemic lupus erythematosus),
CC   respiratory disorders (asthma, allergy), CNS disorders (e.g., Alzheimer's
CC   disease, Parkinson's disease), haematopoietic disorders, cardiovascular
CC   disorders (e.g., atherosclerosis, coronary disease), hyperlipidaemias,
CC   hypolipidaemia, lipidoses, familial hypercholesterolaemia, Gaucher's
CC   disease, Tay-sach's disease, diabetes mellitus, cancer and other enzyme

Query Match            42.3%;     Score 44;   DB 22;   Length 340;
Best Local Similarity  42.9%;     Pred. No. 30;
Matches   9;   Conservative   4;   Mismatches   8;   Indels   0;   Gaps   0;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
          ||  |!:|| ::|| ||:
Db   40 SGYCSTYFRAGSKPFNPVLGE 60


RESULT 6
AAU18677
ID   AAU18677 standard; Protein; 340 AA.
XX
AC   AAU18677;
XX
DT   21-NOV-2001 (first entry)
```

XX
DE Renal and cardiovascular-associated protein, Seq ID 116.
XX
KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
KW cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
KW immunosuppressive; kidney disorder; renal failure; hypertension;
KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
KW hyponatraemia; hyperkalaemia; neoplastic disorder; nephtoma;
KW autoimmune disease; inflammatory disease; reproductive system disorder;
KW endocrine disorder; neural activity; neurological disorder;
KW wound healing; respiratory disorder.
XX
OS Homo sapiens.
XX
PN WO200155328-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01359.
XX
PR 31-JAN-2001; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 23-AUG-2000; 2000US-0227009.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 22-AUG-2000; 2000US-0226681.
PR 06-SEP-2000; 2000US-0230437.
PR 30-AUG-2000; 2000US-0228924.
PR 06-SEP-2000; 2000US-0230438.
PR 01-SEP-2000; 2000US-0229287.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

```
PR    05-DEC-2000;  2000US-0251030.
PR    05-DEC-2000;  2000US-0251988.
PR    05-DEC-2000;  2000US-0256719.
PR    06-DEC-2000;  2000US-0251479.
PR    08-DEC-2000;  2000US-0251856.
PR    08-DEC-2000;  2000US-0251868.
PR    08-DEC-2000;  2000US-0251869.
PR    08-DEC-2000;  2000US-0251989.
PR    08-DEC-2000;  2000US-0251990.
PR    11-DEC-2000;  2000US-0254097.
PR    05-JAN-2001;  2001US-0259678.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-488787/53.
DR    N-PSDB; AAS30198.
XX
PT    New polynucleotides and polypeptides, useful for diagnosing, treating,
PT    preventing or prognosing e.g. kidney, cardiovascular, blood,
PT    electrolyte imbalance or neoplastic disorders, autoimmune diseases,
PT    cancers -
XX
PS    Claim 1; SEQ ID No 116; 506pp; English.
XX
CC    The invention relates to novel nucleic acids and polypeptides useful for
CC    diagnosing, treating, preventing and/or prognosing disorders related to
CC    these polypeptides. The polynucleotides are especially useful in the
CC    diagnosis, prognosis, prevention and/or treatment of diseases which
CC    include kidney disorders (e.g. renal failure or nephritis),
CC    cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC    blood disorders (e.g. anaemia or blood coagulation disorders),
CC    electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC    neoplastic disorders (e.g. nephtoma or renal cell cancer), autoimmune
CC    diseases, cancers, inflammatory diseases, reproductive system
CC    disorders, endocrine disorders, neural activity and neurological
CC    disorders, wound healing and respiratory disorders. AAU18644-AAU18715
CC    represent the novel human renal and cardiovascular-associated amino
CC    acid sequences of the invention. Note: The sequence data for this patent
CC    did not form part of the printed specification, but was obtained in
CC    electronic format directly from WIPO at:
CC    ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence  340 AA;

Query Match             42.3%;   Score 44;  DB 22;  Length 340;
Best Local Similarity   42.9%;   Pred. No. 30;
Matches     9;  Conservative   4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    1  SGSLSTFFRLFNRSFTQALGK 21
          ||  ||:|| ::|:|  ||:
Db   40  SGYCSTYFRAGSKPFNPVLGE 60


RESULT 7
AAU18712
ID    AAU18712 standard; Protein; 340 AA.
XX
AC    AAU18712;
XX
DT    21-NOV-2001  (first entry)
XX
DE    Renal and cardiovascular-associated protein, Seq ID 151.
XX
KW    Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
KW    cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
KW    immunosuppressive; kidney disorder; renal failure; hypertension;
KW    cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
KW    blood coagulation disorder; electrolyte imbalance disorder; cancer;
KW    hyponatraemia; hyperkalaemia; neoplastic disorder; nephtoma;
KW    autoimmune disease; inflammatory disease; reproductive system disorder;
KW    endocrine disorder; neural activity; neurological disorder;
KW    wound healing; respiratory disorder.
XX
OS    Homo sapiens.
XX
PN    WO20015328-A2.
XX
PD    02-AUG-2001.
XX
PF    17-JAN-2001;  2001WO-US01359.
XX
PR    31-JAN-2000;  2000US-0179065.
PR    04-FEB-2000;  2000US-0180628.
PR    24-FEB-2000;  2000US-0184664.
PR    02-MAR-2000;  2000US-0186350.
PR    16-MAR-2000;  2000US-0189874.
PR    17-MAR-2000;  2000US-0190076.
PR    18-APR-2000;  2000US-0198123.
PR    19-MAY-2000;  2000US-0205515.
PR    07-JUN-2000;  2000US-0209467.
PR    28-JUN-2000;  2000US-0214886.
PR    30-JUN-2000;  2000US-0215135.
PR    07-JUL-2000;  2000US-0216647.
PR    07-JUL-2000;  2000US-0216880.
PR    11-JUL-2000;  2000US-0217487.
PR    11-JUL-2000;  2000US-0217496.
PR    14-JUL-2000;  2000US-0218290.
PR    26-JUL-2000;  2000US-0220963.
PR    26-JUL-2000;  2000US-0220964.
PR    14-AUG-2000;  2000US-0224518.
PR    14-AUG-2000;  2000US-0224519.
PR    14-AUG-2000;  2000US-0225213.
PR    14-AUG-2000;  2000US-0225214.
PR    14-AUG-2000;  2000US-0225266.
PR    14-AUG-2000;  2000US-0225267.
PR    14-AUG-2000;  2000US-0225268.
PR    14-AUG-2000;  2000US-0225270.
PR    14-AUG-2000;  2000US-0225447.
PR    14-AUG-2000;  2000US-0225757.
PR    14-AUG-2000;  2000US-0225758.
PR    14-AUG-2000;  2000US-0225759.
PR    14-AUG-2000;  2000US-0226279.
PR    22-AUG-2000;  2000US-0226681.
PR    22-AUG-2000;  2000US-0226868.
PR    23-AUG-2000;  2000US-0227009.
PR    30-AUG-2000;  2000US-0228924.
PR    01-SEP-2000;  2000US-0229287.
PR    01-SEP-2000;  2000US-0229343.
PR    01-SEP-2000;  2000US-0229344.
PR    01-SEP-2000;  2000US-0229345.
PR    05-SEP-2000;  2000US-0229509.
PR    05-SEP-2000;  2000US-0229513.
PR    06-SEP-2000;  2000US-0230437.
PR    06-SEP-2000;  2000US-0230438.
PR    08-SEP-2000;  2000US-0231242.
PR    08-SEP-2000;  2000US-0231243.
PR    08-SEP-2000;  2000US-0231244.
PR    08-SEP-2000;  2000US-0231413.
PR    08-SEP-2000;  2000US-0231414.
PR    08-SEP-2000;  2000US-0232080.
PR    08-SEP-2000;  2000US-0232081.
PR    12-SEP-2000;  2000US-0231968.
PR    14-SEP-2000;  2000US-0232397.
PR    14-SEP-2000;  2000US-0232398.
PR    14-SEP-2000;  2000US-0232399.
PR    14-SEP-2000;  2000US-0232400.
PR    14-SEP-2000;  2000US-0232401.
PR    14-SEP-2000;  2000US-0233063.
PR    14-SEP-2000;  2000US-0233064.
PR    14-SEP-2000;  2000US-0233065.
PR    21-SEP-2000;  2000US-0234223.
PR    21-SEP-2000;  2000US-0234274.
PR    25-SEP-2000;  2000US-0234997.
```

```
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488787/53.
DR   N-PSDB; AAS30233.
XX
PT   New polynucleotides and polypeptides, useful for diagnosing, treating,
PT   preventing or prognosing e.g. kidney, cardiovascular, blood,
PT   electrolyte imbalance or neoplastic disorders, autoimmune diseases,
PT   cancers -
XX
PS   Claim 1; SEQ ID No 151; 506pp; English.
XX
CC   The invention relates to novel nucleic acids and polypeptides useful for
CC   diagnosing, treating, preventing and/or prognosing disorders related to
CC   these polypeptides. The polynucleotides are especially useful in the
CC   diagnosis, prognosis, prevention and/or treatment of diseases which
CC   include kidney disorders (e.g. renal failure or nephritis),
CC   cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC   blood disorders (e.g. anaemia or blood coagulation disorders),
CC   electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC   neoplastic disorders (e.g. nephtoma or renal cell cancer), autoimmune
CC   diseases, cancers, inflammatory diseases, reproductive system
CC   disorders, endocrine disorders, neural activity and neurological
CC   disorders, wound healing and respiratory disorders. AAU18644-AAU18715
CC   represent the novel human renal and cardiovascular-associated amino
CC   acid sequences of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at:
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence    340 AA;

     Query Match             42.3%;  Score 44;  DB 22;  Length 340;
     Best Local Similarity   42.9%;  Pred. No. 30;
     Matches    9;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 SGSLSTFFRLFNRSFTQALGK 21
          || |||:|| .::|| :||:
Db     40 SGYCSTYFRAGSKPFNPVLGE 60


RESULT 8
AAB95255
ID   AAB95255 standard; Protein; 719 AA.
XX
AC   AAB95255;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:17425.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000;  2000EP-0116126.
XX
PR   29-JUL-1999;       99JP-0248036.
PR   27-AUG-1999;       99JP-0300253.
PR   11-JAN-2000;     2000JP-0118776.
PR   02-MAY-2000;     2000JP-0183767.
PR   09-JUN-2000;     2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
```

```
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   06-DEC-2000;   2000US-0255719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2001;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
```

PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI: 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17425; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   719 AA;

Query Match   42.3%; Score 44; DB 22; Length 719;
Best Local Similarity   42.9%; Pred. No. 66;
Matches   9; Conservative   4; Mismatches   8; Indels   0; Gaps   0;

Qy   1 SGSLSTFFRLFNRSFTQALGK 21
          || ||:|| .:|:| || :
Db   419 SGYCSTYFRAGSKPFNPVLGE 439

RESULT 9
AAO07963
ID   AAO07963 standard; Protein; 86 AA.
XX
AC   AAO07963;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 21855.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX

PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI: 2001-514838/56.
DR   N-PSDB; AAI87894.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 21855; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   86 AA;

Query Match   41.3%; Score 43; DB 22; Length 86;
Best Local Similarity   45.0%; Pred. No. 10;
Matches   9; Conservative   5; Mismatches   6; Indels   0; Gaps   0;

Qy   2 GSLSTFFRLFNRSFTQALGK 21
          |:: || .:|| :|| ||| |:
Db   2 GNIHNFFLIGSRSVTQAGGQ 21

RESULT 10
AAW79359
ID   AAW79359 standard; Protein; 106 AA.
XX
AC   AAW79359;
XX
DT   24-NOV-1998 (first entry)
XX
DE   DNA encoding Staphylococcus aureus ssDNA binding protein.
XX
KW   Bacillus subtilis; single-strand binding protein;
KW   helix-destabilising protein; Rattus norvegicus; Eubacterium;
KW   ssDNA binding protein; treatment; prevention; bacterial infection;
KW   Helicobacter pylori; vaccine.
XX
OS   Staphylococcus aureus.
XX
PN   WO9823738-A2.
XX
PD   04-JUN-1998.
XX
PF   24-NOV-1997; 97WO-US22092.
XX
PR   25-NOV-1996; 96US-0031469.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI   Warren RL;
XX
DR   WPI; 1998-322718/28.
DR   N-PSDB; AAV59888.
XX
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
XX

```
PS   Claim 5; Pages 58-59; 114pp; English.
XX
CC   AAW79358-59 represent Staphylococcus aureus WCHU (NCIMB 40771)
CC   proteins that have homology to a Bacillus subtilis single-strand
CC   binding protein (helix-destabilising protein) (AAW79358) and a Rattus
CC   norvegicus and a Eubacterium ssDNA binding protein (AAW79359). The
CC   protein is used to generate antibodies and to screen for antimicrobials.
CC   The products are used to treat or prevent bacterial infections,
CC   particularly where caused by S. aureus but also against Helicobacter
CC   pylori. Particular applications are to treat subjects before surgery or
CC   insertion of an in-dwelling device (alternatively the device itself is
CC   impregnated before placement). The nucleic acid sequence is used as
CC   sources of antisense sequences (for therapeutic use) or regulatory
CC   elements for controlling expression of bacterial genes, and for
CC   antibacterial screening. The protein can be also used as a vaccine.
XX
SQ   Sequence  106 AA;

Query Match             41.3%;  Score 43;  DB 19;  Length 106;
Best Local Similarity   47.4%;             Pred. No. 13;
Matches    9;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 SLSTFFRLFNRSFTQALGK 21
              |:::|| ||:||:|:
Db         25 SVAFTLAVNRTFTNAQGE 43

RESULT 11
AAR41945
ID   AAR41945 standard; Protein; 135 AA.
XX
AC   AAR41945;
XX
DT   10-MAY-1994   (first entry)
XX
DE   Serotonin receptor 5-HT2f C-terminal region.
XX
KW   5-hydroxytryptamine; receptor; serotoninergic; vasoconstriction;
KW   plasmid pHD5HT2f; probe.
XX
PN   EP565370-A.
XX
PD   13-OCT-1993.
XX
PF   07-APR-1993;     93EP-0302759.
PR   09-APR-1992;     92US-0864005.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Baez M,  Kursar JD;
XX
DR   WPI; 1993-322574/41.
DR   N-PSDB; AAQ49783.
XX
PT   5-Hydroxy-tryptamine receptor - used to identify drugs with
PT   receptor activity
XX
PS   Disclosure; Page 17-18; 20pp; English.
XX
CC   The plasmid pHD5HT2f comprises the sequence AAQ49781 which codes for a
CC   novel serotonin receptor. Fragments of the full-length coding
CC   region are disclosed as suitable for use as probes to find
CC   homologous (receptor) sequences. The preferred fragments are those
CC   coding for the G-loop, the N-terminal and the C-terminal of the
CC   5-HT2f receptor (AAR41943-R41945, respectively).
XX
SQ   Sequence  135 AA;

Query Match             41.3%;  Score 43;  DB 14;  Length 135;
Best Local Similarity   38.1%;             Pred. No. 16;
Matches    8;  Conservative    5;  Mismatches    5;  Indels    8;  Gaps    0;
```

```
Qy          1 SGSLSTFFRLFNRSFTQALGK 21
              || |||:||:::|:|| |:
Db         28 SGVNPLIYTLFNKTFREAFGR 48

RESULT 12
AAW79358
ID   AAW79358 standard; Protein; 151 AA.
XX
AC   AAW79358;
XX
DT   24-NOV-1998   (first entry)
XX
DE   DNA encoding Staphylococcus aureus single-strand binding protein.
XX
KW   Bacillus subtilis; single-strand binding protein;
KW   helix-destabilising protein; Rattus norvegicus; Eubacterium;
KW   ssDNA binding protein; treatment; prevention; bacterial infection;
KW   Helicobacter pylori; vaccine.
XX
OS   Staphylococcus aureus.
XX
PN   WO9823738-A2.
XX
PD   04-JUN-1998.
XX
PF   24-NOV-1997;     97WO-US22092.
XX
PR   25-NOV-1996;     96US-0031469.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Warren RL;
XX
DR   WPI; 1998-322718/28.
DR   N-PSDB; AAV59888.
XX
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
XX
PS   Claim 5; Page 58; 114pp; English.
XX
CC   AAW79358-59 represent Staphylococcus aureus WCHU (NCIMB 40771)
CC   proteins that have homology to a Bacillus subtilis single-strand
CC   binding protein (helix-destabilising protein) (AAW79358) and a Rattus
CC   norvegicus and a Eubacterium ssDNA binding protein (AAW79359). The
CC   protein is used to generate antibodies and to screen for antimicrobials.
CC   The products are used to treat or prevent bacterial infections,
CC   particularly where caused by S. aureus but also against Helicobacter
CC   pylori. Particular applications are to treat subjects before surgery or
CC   insertion of an in-dwelling device (alternatively the device itself is
CC   impregnated before placement). The nucleic acid sequence is used as
CC   sources of antisense sequences (for therapeutic use) or regulatory
CC   elements for controlling expression of bacterial genes, and for
CC   antibacterial screening. The protein can be also used as a vaccine.
XX
SQ   Sequence  151 AA;

Query Match             41.3%;  Score 43;  DB 19;  Length 151;
Best Local Similarity   47.4%;             Pred. No. 18;
Matches    9;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 SLSTFFRLFNRSFTQALGK 21
              |:::|| ||:||:|:
Db         25 SVAFTLAVNRTFTNAQGE 43

RESULT 13
AAU34248
ID   AAU34248 standard; Protein; 167 AA.
XX
AC   AAU34248;
```

```
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #524.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52107.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5744; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   167 AA;

Query Match      41.3%;  Score 43;  DB 22;  Length 167;
Best Local Similarity   47.4%;  Pred. No. 21;
Matches   9;  Conservative   4;  Mismatches   4;  Indels   6;  Gaps   0;

Qy    3  SLSTFFRLFNRSFTQALGK 21
          |:::|| |:::|| |:
Db   25  SVAIFTLAVNRTFTNAQGE 43


RESULT 14
AAU36819
ID   AAU36819 standard; Protein; 167 AA.
XX
AC   AAU36819;
```

```
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #989.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54678.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12412; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   167 AA;

Query Match      41.3%;  Score 43;  DB 22;  Length 167;
Best Local Similarity   47.4%;  Pred. No. 21;
Matches   9;  Conservative   4;  Mismatches   4;  Indels   6;  Gaps   0;

Qy    3  SLSTFFRLFNRSFTQALGK 21
          |:::|| |:::|| |:
Db   25  SVAIFTLAVNRTFTNAQGE 43


RESULT 15
AAU37429
ID   AAU37429 standard; Protein; 167 AA.
XX
AC   AAU37429;
```

```
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1599.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT, Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS55288.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 13022; 51pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  167 AA;

Query Match        41.3%;  Score 43;  DB 22;  Length 167;
Best Local Similarity   47.4%;  Pred. No. 21;
Matches    9;  Conservative   4;  Mismatches   6;  Indels   0;  Gaps   0;

Qy      3  SLSTFPRLFNRSFTQALGK 21
           |:::|   |!:!|  | |:
Db     25  SVATFTLAVNRTFTNAQGE 43
```

Search completed: November 5, 2002, 10:56:08
Job time : 18.9254 secs

OM protein - protein search, using sw model

```
              GenCore version 5.1.3
        Copyright (c) 1993 - 2002 Compugen Ltd.
```

Run on: November 5, 2002, 10:55:03 ; Search time 6.58209 Seconds
                                 (without alignments)
                                 77.929 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 104
Sequence: 1 SGGLSTPFRLFNRSFTQALGK 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
        1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
        2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
        3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
        4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
        5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
        6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	43	41.3	479	1	US-08-416-788-2	Sequence 2, Appli
2	41	39.4	437	1	US-08-764-343-1	Sequence 1, Appli
3	41	39.4	437	1	US-08-989-925-3	Sequence 3, Appli
4	40	38.5	82	4	US-08-482-918-59	Sequence 59, Appl
5	40	38.5	82	4	US-09-224-681-59	Sequence 59, Appl
6	40	38.5	82	4	US-08-336-728A-59	Sequence 59, Appl
7	40	38.5	135	4	US-08-936-165A-348	Sequence 348, App
8	40	38.5	164	1	US-08-318-193-77	Sequence 77, Appl
9	40	38.5	165	2	US-08-955-848A-1	Sequence 1, Appli
10	40	38.5	165	5	PCT-US95-03866-2	Sequence 2, Appli
11	40	38.5	166	2	US-08-628-428-2	Sequence 2, Appli
12	40	38.5	166	2	US-08-628-428-5	Sequence 5, Appli
13	40	38.5	166	2	US-08-628-428-8	Sequence 8, Appli
14	40	38.5	166	4	US-09-106-891-2	Sequence 2, Appli
15	40	38.5	166	4	US-08-172-507-2	Sequence 2, Appli
16	40	38.5	166	5	PCT-US95-03866-18	Sequence 18, Appl
17	40	38.5	166	5	PCT-US95-03866-34	Sequence 34, Appl
18	40	38.5	166	5	PCT-US95-03866-36	Sequence 36, Appl
19	40	38.5	167	5	PCT-US95-03866-20	Sequence 20, Appl
20	40	38.5	169	2	US-08-955-848A-2	Sequence 2, Appli
21	40	38.5	169	2	US-08-955-848A-3	Sequence 3, Appli
22	40	38.5	169	2	US-08-955-848A-4	Sequence 4, Appli
23	40	38.5	169	2	US-08-955-848A-5	Sequence 5, Appli
24	40	38.5	169	2	US-08-955-848A-6	Sequence 6, Appli
25	40	38.5	169	2	US-08-955-848A-7	Sequence 7, Appli
26	40	38.5	169	2	US-08-955-848A-8	Sequence 8, Appli
27	40	38.5	169	2	US-08-955-848A-9	Sequence 9, Appli
28	40	38.5	169	2	US-08-955-848A-10	Sequence 10, Appl
29	40	38.5	169	2	US-08-955-848A-11	Sequence 11, Appl
30	40	38.5	169	2	US-08-955-848A-12	Sequence 12, Appl
31	40	38.5	169	2	US-08-955-848A-13	Sequence 13, Appl
32	40	38.5	169	2	US-08-955-848A-14	Sequence 14, Appl
33	40	38.5	169	2	US-08-955-848A-15	Sequence 15, Appl
34	40	38.5	169	2	US-08-955-848A-16	Sequence 16, Appl
35	40	38.5	169	2	US-08-955-848A-17	Sequence 17, Appl
36	40	38.5	169	2	US-08-955-848A-18	Sequence 18, Appl
37	40	38.5	169	2	US-08-955-848A-19	Sequence 19, Appl
38	40	38.5	169	2	US-08-955-848A-20	Sequence 20, Appl
39	40	38.5	169	2	US-08-955-848A-21	Sequence 21, Appl
40	40	38.5	169	2	US-08-955-848A-22	Sequence 22, Appl
41	40	38.5	169	2	US-08-955-848A-23	Sequence 23, Appl
42	40	38.5	169	2	US-08-955-848A-24	Sequence 24, Appl
43	40	38.5	169	2	US-08-955-848A-25	Sequence 25, Appl
44	40	38.5	169	2	US-08-955-848A-26	Sequence 26, Appl
45	40	38.5	169	2	US-08-955-848A-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-416-788-2
; Sequence 2, Application US/08416788
; Patent No. 5780245
; GENERAL INFORMATION:
; APPLICANT: Maroteaux, Luc
; TITLE OF INVENTION: No. 5780245el Polypeptides Having a Serotonin
; TITLE OF INVENTION: Receptor Activity, Nucleic Acids Coding for These
; TITLE OF INVENTION: Polypeptides and Uses
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Rhone-Poulenc Rorer Inc.
;   STREET: 500 Arcola Road, 3C43
;   CITY: Collegeville
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19426-0107
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/416,788
;   FILING DATE:
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/FR93/01012
;   FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: FR 92-12280
;   FILING DATE: 14-OCT-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, Julie K.
;   REGISTRATION NUMBER: 38,619
;   REFERENCE/DOCKET NUMBER: EX92008-US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (610)454-3839
;   TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 479 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-416-788-2

  Query Match   41.3%; Score 43; DB 1; Length 479;
  Best Local Similarity   38.1%; Pred. No. 12;
  Matches   8; Conservative   8; Mismatches   5; Indels   0; Gaps   0;

```
Qy    1 SGSLSTFFRLFNRSFTQALGK 21
         ||  :| :|||:|| :|| :|:
Db  372 SGVNPLIYTLFNKTFREAFGR 392
```

RESULT 2
US-08-764-343-1
; Sequence 1, Application US/08764343
; Patent No. 5739009
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Hawkins, Phillip R.
;   TITLE OF INVENTION: NOVEL ADIPOCYTE-SPECIFIC
;   TITLE OF INVENTION: DIFFERENTIATION-RELATED PROTEIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: US
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/764,343
;   FILING DATE: Filed Herewith
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0167 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 437 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   IMMEDIATE SOURCE:
;   CLONE: Consensus
;US-08-764-343-1

Query Match   39.4%; Score 41; DB 1; Length 437;
Best Local Similarity   50.0%; Pred. No. 24;
Matches   10; Conservative   5; Mismatches   5; Indels   3; Gaps   2;

```
Qy    2 GSLSTFFRLFNRSFTQALGK 21
        |||||  :| :|:: ||| :|||
Db  221 GSLST--KLHSRAYQQALSR 238
```

RESULT 3
US-08-989-925-3
; Sequence 3, Application US/08989925
; Patent No. 5989820
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Lal, Preeti
;   APPLICANT: Shah, Purvi
;   APPLICANT: Corley, Neil C.
;   TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/989,925
;   FILING DATE: Herewith
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0440 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650-855-0555
;   TELEFAX: 650-845-4166
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 437 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY: GenBank
;   CLONE: 1806040
;US-08-989-925-3

Query Match   39.4%; Score 41; DB 2; Length 437;
Best Local Similarity   50.0%; Pred. No. 24;
Matches   10; Conservative   5; Mismatches   5; Indels   3; Gaps   1;

```
Qy    2 GSLSTFFRLFNRSFTQALGK 21
        |||||  :| :|:: |||
Db  221 GSLST--KLHSRAYQQALSR 238
```

RESULT 4
US-08-482-918-59
; Sequence 59, Application US/08482918
; Patent No. 6207417
; GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Stem Cell Factor
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/482,918
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/33005
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 59:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 82 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-482-918-59

Query Match   38.5%;   Score 40;   DB 4;   Length 82;
Best Local Similarity   87.5%;   Pred. No. 5.1;
Matches   7;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   7 FFRLFNRS 14
        ||:||||
Db   2 FFRIFNRS 9

RESULT 5
US-09-224-681-59
; Sequence 59, Application US/09224681
; Patent No. 6207454
GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Method for Enhancing the Efficiency of Gene
;   TITLE OF INVENTION: Transfer with Stem Cell Factor (SCF) Polypeptide
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;   ADDRESSE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/224,681
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/005,893
;   FILING DATE: 12-JAN-1998
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/449,653
;   FILING DATE: 24-MAY-1995
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/589,701
;   FILING DATE: 01-OCT-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/573,616
;   FILING DATE: 24-AUG-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/537,198
;   FILING DATE: 11-JUN-1990

;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/422,383
;   FILING DATE: 16-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/35199
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX:
; INFORMATION FOR SEQ ID NO: 59:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 82 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-224-681-59

Query Match   38.5%;   Score 40;   DB 4;   Length 82;
Best Local Similarity   87.5%;   Pred. No. 5.1;
Matches   7;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   7 FFRLFNRS 14
        ||:||||
Db   2 FFRIFNRS 9

RESULT 6
US-08-336-728A-59
; Sequence 59, Application US/08336728A
; Patent No. 6207802
GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Stem Cell Factor
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;   ADDRESSE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/336,728A
;   FILING DATE: 09-NOV-1994
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/982,255
;   FILING DATE: 25-NOV-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/589,701
;   FILING DATE: 01-OCT-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/573,616
;   FILING DATE: 24-AUG-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/537,198
;   FILING DATE: 11-JUN-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/422,383
;   FILING DATE: 16-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107

```
;   REFERENCE/DOCKET NUMBER: 01017/32956
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 312/474-6300
;     TELEFAX: 312/474-0448
;     TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 59:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 82 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-336-728A-59

Query Match             38.5%;   Score 40;  DB 4;   Length 82;
Best Local Similarity   87.5%;   Pred. No. 5.1;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    7  FFRLFNRS 14
         |||:||||
Db    2  FFRIFNRS 9

RESULT 7
US-08-936-165A-348
; Sequence 348, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
;     APPLICANT: Black, Michael
;     APPLICANT: Burnham, Martin
;     APPLICANT: Hodgson, John
;     APPLICANT: Knowles, David
;     APPLICANT: Lonetto, Michael
;     APPLICANT: Nicholas, Richard
;     APPLICANT: Pratt, Julie
;     APPLICANT: Reichard, Richard
;     APPLICANT: Rosenberg, Martin
;     APPLICANT: Ward, Judith
;     TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
;     TITLE OF INVENTION: Polypeptides and Their Uses
;     NUMBER OF SEQUENCES: 534
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE: SmithKline Beecham Corporation
;         STREET: 709 Swedeland Road
;         CITY: King of Prussia
;         STATE: PA
;         COUNTRY: USA
;         ZIP: 19406-0939
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE: Diskette
;         COMPUTER: IBM Compatible
;         OPERATING SYSTEM: DOS
;         SOFTWARE: FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/08/936,165A
;         FILING DATE: 24-SEP-1997
;         CLASSIFICATION: 536
;     PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: 60/027,032
;         FILING DATE: 24-SEP-1996
;     ATTORNEY/AGENT INFORMATION:
;         NAME: Gimmi, Edward R
;         REGISTRATION NUMBER: 38,891
;         REFERENCE/DOCKET NUMBER: P50549
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 610-270-4478
;         TELEFAX: 610-270-5090
;         TELEX:
; INFORMATION FOR SEQ ID NO: 348:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 135 amino acids
;         TYPE: amino acid
;         STRANDEDNESS: single
;         TOPOLOGY: linear
;     MOLECULE TYPE: Protein
US-08-936-165A-348

Query Match             38.5%;   Score 40;  DB 4;   Length 135;
Best Local Similarity   47.4%;   Pred. No. 9.1;
Matches    9;  Conservative    3;  Mismatches    3;  Indels    7;  Gaps    0;

Qy    3  SLSTFFRLFNRSFTQALGK 21
         |:||  ||:||:|  |:
Db   19  SXTTFTIAVNRTFTNAQGE 37

RESULT 8
US-08-318-193-77
; Sequence 77, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
;     APPLICANT: GARVIN, Robert T.
;     APPLICANT: MALEK, Lawrence T.
;     TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
;     TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
;     TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
;     TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
;     NUMBER OF SEQUENCES: 91
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Foley & Lardner
;         STREET: 1800 Diagonal Road, Suite 500
;         CITY: Alexandria
;         STATE: Virginia
;         COUNTRY: USA
;         ZIP: 22313-0299
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/08/318,193
;         FILING DATE:
;         CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: US/07/935,314
;         FILING DATE:
;         APPLICATION NUMBER: US 07/224,568
;     ATTORNEY/AGENT INFORMATION:
;         NAME: BENT, Stephen A.
;         REGISTRATION NUMBER: 29,768
;         REFERENCE/DOCKET NUMBER: 18740/116 CACO
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (703)836-9300
;         TELEFAX: (703)683-4109
;         TELEX: 899149
; INFORMATION FOR SEQ ID NO: 77:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 164 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-318-193-77

Query Match             38.5%;   Score 40;  DB 1;   Length 164;
Best Local Similarity   87.5%;   Pred. No. 11;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    7  FFRLFNRS 14
         |||:||||
Db  115  FFRIFNRS 122

RESULT 9
US-08-955-848A-1
; Sequence 1, Application US/08955848A
; Patent No. 5969105
```

```
  GENERAL INFORMATION:
  APPLICANT: Mc Wherter, Charles
  APPLICANT: Feng, Yiqing
  TITLE OF INVENTION: No. 5969105e1 Stem Cell Factor Receptor
  TITLE OF INVENTION: Agonists
  NUMBER OF SEQUENCES: 86
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: G. D. Searle & Co.
  STREET: P.O. Box 5110
  CITY: Chicago
  STATE: IL
  COUNTRY: U. S. A.
  ZIP: 60680
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/955,848A
  FILING DATE: 21-OCT-1997
  CLASSIFICATION: 536
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 60/029,165
  FILING DATE: 25-OCT-1997
  ATTORNEY/AGENT INFORMATION:
  NAME: Bennett, Dennis A
  REGISTRATION NUMBER: 34,547
  REFERENCE/DOCKET NUMBER: C-2992/1
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 314-737-6986
  TELEFAX: 314-737-6972
  TELEX:
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 165 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
US-08-955-848A-1

Query Match     38.5%;    Score 40;  DB 2;  Length 165;
Best Local Similarity  87.5%;    Pred. No. 11;
Matches  7;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    7  FFRLFNRS 14
         |||:|||
Db  115  FFRIFNRS 122

RESULT 10
PCT-US95-03866-2
; Sequence 2, Application PC/TUS9503866
; GENERAL INFORMATION:
; APPLICANT: Cytomed, Inc. (all states except US)
; APPLICANT: Nocka, Karl (US only)
; APPLICANT: Lobell, Robert B (US only)
; TITLE OF INVENTION: STABILIZED DIMER OF KIT LIGAND AND
; TITLE OF INVENTION: FLT-3/FLK-2 LIGAND
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/03866
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/220,379
;   FILING DATE: 28-MAR-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Haley Jr, James F
;   REGISTRATION NUMBER: 27,794
;   REFERENCE/DOCKET NUMBER: CytoMed/2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-596-9000
;   TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 165 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-03866-2

Query Match     38.5%;    Score 40;  DB 5;  Length 165;
Best Local Similarity  87.5%;    Pred. No. 11;
Matches  7;  Conservative  1;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    7  FFRLFNRS 14
         |||:|||
Db  115  FFRIFNRS 122

RESULT 11
US-08-628-428-2
; Sequence 2, Application US/08628428
; Patent No. 5885962
; GENERAL INFORMATION:
; APPLICANT: Lu, Hsieng
; TITLE OF INVENTION: SCF ANALOG COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amgen Inc.
;   STREET: 1840 DeHavilland Drive
;   CITY: Thousand Oaks
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 91320-1789
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/628,428
;   FILING DATE: 05-APR-1996
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Knight, Matthew W
;   REGISTRATION NUMBER: 36,846
;   REFERENCE/DOCKET NUMBER: A-400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 166 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
;   NAME/KEY: Protein
;   LOCATION: 1..166
;   OTHER INFORMATION: /note= "Met sequence starts at -1
;   OTHER INFORMATION: on Sequence No. 5885962 2."
; Patent No. 5885962
US-08-628-428-2
```

Query Match   38.5%;   Score 40; DB 2; Length 166;
Best Local Similarity 87.5%;   Pred. No. 12;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   7 FFRLFNRS 14
          |||:||||
Db   116 FFRIFNRS 123

RESULT 12
US-08-628-428-5
; Sequence 5, Application US/08628428
; Patent No. 5885962
; GENERAL INFORMATION:
;   APPLICANT: Lu, Hsieng
;   TITLE OF INVENTION: SCF ANALOG COMPOSITIONS AND METHODS
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amgen Inc.
;   STREET: 1840 DeHavilland Drive
;   CITY: Thousand Oaks
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 91320-1789
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/628,428
;   FILING DATE: 05-APR-1996
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Knight, Matthew W
;   REGISTRATION NUMBER: 36,846
;   REFERENCE/DOCKET NUMBER: A-400
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 166 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;   NAME/KEY: Protein
;   LOCATION: 1..166
;   OTHER INFORMATION: /note= "Met sequence starts at -1
;   OTHER INFORMATION: on Sequence NO. 5885962 5."
; Patent No. 5885962
US-08-628-428-5

Query Match   38.5%;   Score 40; DB 2; Length 166;
Best Local Similarity 87.5%;   Pred. No. 12;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   7 FFRLFNRS 14
          |||:||||
Db   116 FFRIFNRS 123

RESULT 13
US-08-628-428-8
; Sequence 8, Application US/08628428
; Patent No. 5885962
; GENERAL INFORMATION:
;   APPLICANT: Lu, Hsieng
;   TITLE OF INVENTION: SCF ANALOG COMPOSITIONS AND METHODS
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amgen Inc.
;   STREET: 1840 DeHavilland Drive
;   CITY: Thousand Oaks
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 91320-1789
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/628,428
;   FILING DATE: 05-APR-1996
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Knight, Matthew W
;   REGISTRATION NUMBER: 36,846
;   REFERENCE/DOCKET NUMBER: A-400
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 166 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;   NAME/KEY: Protein
;   LOCATION: 1..166
;   OTHER INFORMATION: /note= "Met sequence starts at -1
;   OTHER INFORMATION: on Sequence NO. 5885962 8."
; Patent No. 5885962
US-08-628-428-8

Query Match   38.5%;   Score 40; DB 2; Length 166;
Best Local Similarity 87.5%;   Pred. No. 12;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   7 FFRLFNRS 14
          |||:||||
Db   116 FFRIFNRS 123

RESULT 14
US-09-106-891-2
; Sequence 2, Application US/09106891
; Patent No. 5965522
; GENERAL INFORMATION:
;   APPLICANT: Hershenson, Susan
;   TITLE OF INVENTION: No. 5965522el Stem Cell Factor Formulations and
;   TITLE OF INVENTION: Methods
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amgen Inc.
;   STREET: 1840 DeHavilland Drive
;   CITY: Thousand Oaks
;   STATE: California
;   COUNTRY: USA
;   ZIP: 91320
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/106,891
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/172,507
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pessin, Karol M.
;   REFERENCE/DOCKET NUMBER: A-276
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  166 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-106-891-2

Query Match             38.5%;  Score 40;  DB 2;  Length 166;
Best Local Similarity   87.5%;  Pred. No. 12;
Matches     7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 FFRLFNRS 14
            |||:||||
Db     116 FFRIFNRS 123


RESULT 15
US-08-172-507-2
; Sequence 2, Application US/08172507
; Patent No. 6288030
; GENERAL INFORMATION:
;      APPLICANT:  Hershenson, Susan
;      TITLE OF INVENTION:  No. 6288030el Stem Cell Factor Formulations and
;      TITLE OF INVENTION:  Methods
;      NUMBER OF SEQUENCES:  2
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Amgen Inc.
;        STREET:  1840 Dehavilland Drive
;        CITY:  Thousand Oaks
;        STATE:  California
;        COUNTRY:  USA
;        ZIP:  91320
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC Compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/172,507
;        FILING DATE:
;        CLASSIFICATION:  514
;      ATTORNEY/AGENT INFORMATION:
;        NAME:  Pessin, Karol M.
;        REFERENCE/DOCKET NUMBER:  A-276
; INFORMATION FOR SEQ ID NO:  2:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  166 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-172-507-2

Query Match             38.5%;  Score 40;  DB 4;  Length 166;
Best Local Similarity   87.5%;  Pred. No. 12;
Matches     7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 FFRLFNRS 14
            |||:||||
Db     116 FFRIFNRS 123


Search completed: November  5, 2002, 10:58:19
Job time : 6.58209 secs
```

```
                 GenCore version 5.1.3
        Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     November  5, 2002, 10:55:02 ; Search time 8.14925 Seconds
                                          (without alignments)
                                          247.615 Million cell updates/sec

Title:       US-09-833-017B-4
Perfect score:  104
Sequence:    1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:    283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters:     283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_71:*
                1: pir1:*
                2: pir2:*
                3: pir3:*
                4: pir4:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.


                                 SUMMARIES

                %
Result          Query
No.   Score  Match  Length DB  ID       Description
----  -----  -----  ------ --  ------   -----------
1      48    46.2   1083   2   S76111   acriflavin resista
2      45    43.3   124    2   I51577   gene wnt-7C protei
3      45    43.3   850    2   F95260   ABC transporter, p
4      45    43.3   850    1   A98126   conserved hypothet
5      44.5  42.8   689    1   S29621   UDPglucose 4-epime
6      44    42.3   490    2   I48163   cytochrome P450 -
7      44    42.3   889    2   H84506   probable retroelem
8      44    42.3   970    2   G84939   hypothetical prote
9      43    41.3   167    2   G89802   hypothetical prote
10     43    41.3   230    2   S46737   hypothetical prote
11     43    41.3   305    2   F67748   phosphoprotein pho
12     43    41.3   412    2   C38351   phosphoprotein pho
13     43    41.3   443    2   B38351   protein phosphatas
14     43    41.3   443    2   S65685   phosphoprotein pho
15     43    41.3   447    2   A38351   phosphoprotein pho
16     43    41.3   447    2   A41805   phosphoprotein pho
17     43    41.3   468    2   JC5417   phosphoprotein pho
18     43    41.3   468    2   S65951   [phosphorylase] ph
19     43    41.3   479    2   S23562   serotonin receptor
20     43    41.3   504    2   S27269   serotonin receptor
21     43    41.3   835    1   P3XRSR   minor inner core p
22     43    41.3   1123   2   D36790   hypothetical prote
23     42.5  40.9   361    2   D98243   opuaa (AF234619) [
24     42.5  40.9   361    1   AH3042   hypothetical prote
25     42    40.4   113    2   T09627   positive transcrip
26     42    40.4   139    2   A71190   hypothetical prote
27     42    40.4   155    2   T31981   hypothetical prote
28     42    40.4   206    2   A81231   50S ribosomal prot
29     42    40.4   393    2   E82283   conserved hypothet
```

```
30   42    490    2   I48162   cytochrome P450 -
31   42    490    2   I48189   cytochrome P450 II
32   42    499    2   A45778   phosphoprotein pho
33   42    709    2   S73497   cell division prot
34   42    851    2   A83484   probable heme util
35   41.5  1242   2   S51246   probable DNA repai
36   41    136    2   T00295   probable transposa
37   41    158    2   D71111   hypothetical prote
38   41    253    2   S70367   stem cell factor s
39   41    262    2   F83546   probable transcrip
40   41    265    1   BVECIT   istB protein - Esc
41   41    287    2   S70366   stem cell factor 1
42   41    373    2   T47150   hypothetical prote
43   41    435    2   E82848   alpha-ketoglutarat
44   41    527    2   T39927   probable serine-ri
45   41    556    2   T39476   probable serine-ri
```

40.4 (column values for lines 30-45)

```
                              ALIGNMENTS

RESULT 1
S76111
acriflavin resistance protein envD - Synechocystis sp.
N;Alternate names: protein slr0369
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76111
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76111
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1083 <KAN>
A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; GB:AB001339; NID:g1001396; PIDN:BAA10089.1; PID:g100
C;Genetics:
A;Gene: envD
C;Superfamily: acriflavin resistance protein

   Query Match          46.2%;       Score 48; DB 2;  Length 1083;
   Best Local Similarity  64.3%;     Pred. No. 10;
   Matches    9;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy       2  GSLSTFFRLFNRSF  15
            |:|:|| ||||:|
Db     514  GPLAWFFNLFNRTF  527


RESULT 2
I51577
gene wnt-7C protein - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
C;Accession: I51577
R;Wolda, S.L.; Moon, R.T.
Oncogene 7, 1941-1947, 1992
A;Title: Cloning and developmental expression in Xenopus laevis of seven additional m
A;Reference number: I51571; MUID:93026368
A;Accession: I51577
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-124 <WOL>
A;Cross-references: GB:L07535; NID:g214889; PIDN:AAA49988.1; PID:g214890
C;Genetics:
A;Gene: wnt-7C
C;Superfamily: int-1 transforming protein
```

```
Query Match     43.3%;    Score 45; DB 2;  Length 124;
Best Local Similarity 50.0%;  Pred. No. 3.3;
Matches    9;  Conservative  4;  Mismatches  5;  Indels  0;  Gaps  0;

Qy    1 SGSLSTFFRLFNRSFTQA 18
        ||:||:  ||:|||
Db   81 TGSIGTYGRFCNRTSTQA 98
```

RESULT 3
F95260
ABC transporter, permease protein, probable SP2231 [imported] - Streptococcus pneumoniae
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: F95260
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: F95260
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-850 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK76279.1; PID:g14973742; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP2231

```
Query Match     43.3%;    Score 45; DB 2;  Length 850;
Best Local Similarity 50.0%;  Pred. No. 25;
Matches    8;  Conservative  8;  Mismatches  5;  Indels  3;  Gaps  0;

Qy    6 TFFRLFNRSFTQALGK 21
        ||:||  ::||:|:| |
Db  733 TFYRLDTKTFTEAIQK 748
```

RESULT 4
A98126
conserved hypothetical protein spr2036 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: A98126
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A98126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-850 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00838.1; PID:g15459743; GSPDB:GN00174
C;Genetics:
A;Gene: spr2036

```
Query Match     43.3%;    Score 45; DB 2;  Length 850;
Best Local Similarity 50.0%;  Pred. No. 25;
Matches    8;  Conservative  8;  Mismatches  5;  Indels  3;  Gaps  0;

Qy    6 TFFRLFNRSFTQALGK 21
        ||:||  ::||:|:| |
Db  733 TFYRLDTKTFTEAIQK 748
```

RESULT 5
S29621
UDPglucose 4-epimerase (EC 5.1.3.2) - yeast (Pachysolen tannophilus)

N:Alternate names: UDPgalactose 4-epimerase
C;Species: Pachysolen tannophilus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S29621
R:Skrzypek, M.; Maleszka, R.
submitted to the EMBL Data Library, October 1992
A;Description: Cloning and sequencing of the UDP-galactose-4-epimerase gene from Pach
A;Reference number: S29621
A;Accession: S29621
A;Molecule type: DNA
A;Residues: 1-689 <SKR>
A;Cross-references: EMBL:X68593; NID:g3264; PIDN:CAA48580.1; PID:g3265
C;Genetics:
A;Gene: GAL10
C;Superfamily: yeast UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
C;Keywords: galactose metabolism; isomerase
F:4-343/Domain: UDPglucose 4-epimerase homology <UDP>

```
Query Match     42.8%;    Score 44.5; DB 1;  Length 689;
Best Local Similarity 47.6%;  Pred. No. 24;
Matches   10;  Conservative  6;  Mismatches  4;  Indels  1;  Gaps  1;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
        :|::|||: ||::|:   :|:|:||
Db  270 TGNGSTVFEVFN-AFCEAVGK 289
```

RESULT 6
I48163
cytochrome P450 - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Mar-2000
C;Accession: I48163
R:Sakuma, T.; Masaki, K.; Itoh, S.; Yokoi, T.; Kamataki, T.
Mol. Pharmacol. 45, 228-236, 1994
A;Title: Sex-related difference in the expression of cytochrome P450 in hamsters: CDN
A;Reference number: I48162; MUID:9158799
A;Accession: I48163
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-490 <RES>
A;Cross-references: GB:D11436; NID:g220312; PIDN:BAA02002.1; PID:g220313
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; transmembrane protein
F:294-457/Domain: cytochrome P450 homology <P45>
F:435/Binding site: heme iron (Cys) (axial ligand) #status predicted

```
Query Match     42.3%;    Score 44; DB 2;  Length 490;
Best Local Similarity 42.9%;  Pred. No. 21;
Matches    9;  Conservative  5;  Mismatches  7;  Indels  0;  Gaps  0;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
        ||||: :  ||
Db   51 SGSLTNFSKVYGPVFTLYLGR 71
```

RESULT 7
H84506
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84506
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: H84506
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-889 <STO>
A;Cross-references: GB:AE002093; NID:g4417309; PIDN:AAD20433.1; GSPDB:GN00139

C;Genetics:
A;Gene: At2g13330
A;Map position: 2

Query Match   42.3%; Score 44; DB 2; Length 889;
Best Local Similarity 52.9%; Pred. No. 38;
Matches   9; Conservative   2; Mismatches   6; Indels   0; Gaps   0;

Qy   5 STFFRLFNRSFTQALGK 21
           :|:|| || ||| |||
Db   375 TTYQRLVNRMFVDQLGK 391

RESULT 8
G84939
hypothetical protein ytfN [imported] - Buchnera sp. (strain APS)
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: G84939
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
Nature 407, 81-86, 2000
A;Reference number: A84930; MUID:20445173
A;Accession: G84939
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-970 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
C;Genetics:
A;Experimental source: strain APS
A;Gene: ytfN; BU087

Query Match   42.3%; Score 44; DB 2; Length 970;
Best Local Similarity 80.0%; Pred. No. 42;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   6 TFFRLFNRSF 15
           :|:||| |||||
Db   703 SFFNLFNRSF 712

RESULT 9
G89802
hypothetical protein ssb [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: G89802
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kanamoto, Y.; Kobayashi, N.; Sawano, T.; Inoue, T.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: G89802
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-167 <KUR>
A;Cross-references: GB:BA000018; PID:g13700280; PIDN:BAB41578.1; GSPDB:GN00149
C;Genetics:
A;Experimental source: strain N315
A;Gene: ssb
C;Superfamily: bacterial single-stranded DNA-binding protein; single-stranded DNA-bindin

Query Match   41.3%; Score 43; DB 2; Length 167;
Best Local Similarity 47.4%; Pred. No. 9.7;
Matches   9; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   3 SLSTFFRLFNRSFTQALGK 21
           |:|| ||:||| | :
Db   25 SVATFTLAVNRTFTNAQGE 43

RESULT 10

S46737
hypothetical protein YHR038w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein H8179.10|
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C;Accession: S46737
R;Du, Z.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 8179.
A;Reference number: S46732
A;Accession: S46737
A;Molecule type: DNA
A;Residues: 1-230 <DU2>
A;Cross-references: EMBL:U00062; NID:g488162; PID:g488171; GSPDB:GN00008; MIPS:YHR038
C;Genetics:
A;Gene: MIPS:YHR038w
A;Map position: 8R

Query Match   41.3%; Score 43; DB 2; Length 230;
Best Local Similarity 80.0%; Pred. No. 14;
Matches   8; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   9 RLFNRSFTQA 18
           ||||||||:|:
Db   17 RLFNRSFSQS 26

RESULT 11
F69748
hypothetical protein ybfA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: F69748
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: F69748
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-305 <KUN>
A;Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12010.1; PIDN:e11821
A;Experimental source: strain 168
C;Genetics:
A;Gene: ybfA

Query Match   41.3%; Score 43; DB 2; Length 305;
Best Local Similarity 50.0%; Pred. No. 18;
Matches   9; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   3 SLSTFFRLFNRSFTQALG 20
           || ||| |:|| |
Db   7 SVGLKFKFNRFYFNVLG 24

RESULT 12
C38351
phosphoprotein phosphatase (EC 3.1.3.16) 2A-beta 55K regulatory chain - rabbit (fragm
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 23-Jun-1993
C;Accession: C38351

R:Mayer, R.E.; Hendrix, P.; Cron, P.; Matthies, R.; Stone, S.R.; Goris, J.; Merlevede, W
Biochemistry 30, 3589-3597, 1991
A;Title: Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: evidence
A;Reference number: A38351; MUID:9119806
A;Accession: C38351
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-412 <MAY>
A;Cross-references: GB:J05328
C;Keywords: phosphoric monoester hydrolase

```
            Query Match          41.3%;    Score 43;  DB  2;  Length 412;
            Best Local Similarity 50.0%;   Pred. No. 25;
            Matches    7;  Conservative  5;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1  SGSLSTFFRLFNRS 14
            :||.: |||:|:|::
Db    324  TGSYNNFFRMFDRN 337
```

RESULT 13
B38351
phosphoprotein phosphatase 2-beta regulatory chain - human
N;Alternate names: phosphoprotein phosphatase 2A-beta 55K regulatory chain B
C;Species: Homo sapiens (man)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 05-Nov-1999
C;Accession: B38351
R:Mayer, R.E.; Hendrix, P.; Cron, P.; Matthies, R.; Stone, S.R.; Goris, J.; Merlevede, W
Biochemistry 30, 3589-3597, 1991
A;Title: Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: evidence
A;Reference number: A38351; MUID:9119806
A;Accession: B38351
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-443 <MAY>
A;Cross-references: GB:M64930; GB:J05328; NID:g190423; PIDN:AAA36493.1; PID:g190426
C;Genetics:
A;Gene: GDB:PPP2R2B
A;Cross-references: GDB:136801

```
            Query Match          41.3%;    Score 43;  DB  2;  Length 443;
            Best Local Similarity 50.0%;   Pred. No. 27;
            Matches    7;  Conservative  5;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1  SGSLSTFFRLFNRS 14
            :||.: |||:|:|:
Db    355  TGSYNNFFRMFDRN 368
```

RESULT 14
S65685
protein phosphatase 2A, chain B (BR beta) - rat
C;Species: Rattus sp. (rat)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C;Accession: S65685
R:Akiyama, N.; Shima, H.; Hatano, Y.; Osawa, Y.; Sugimura, T.; Nagao, M.
Eur. J. Biochem. 230, 766-772, 1995
A;Title: cDNA cloning of BR-gamma, a novel brain-specific isoform of the B regulatory su
A;Reference number: S65685; MUID:9533116
A;Accession: S65685
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-443 <AKI>
A;Cross-references: EMBL:D38260; NID:g1065605; PIDN:BAA07412.1; PID:d1007991; PID:g17773

```
            Query Match          41.3%;    Score 43;  DB  2;  Length 443;
            Best Local Similarity 50.0%;   Pred. No. 27;
            Matches    7;  Conservative  5;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1  SGSLSTFFRLFNRS 14
            :||.: |||:|:|:
Db    355  TGSYNNFFRMFDRN 368
```

RESULT 15
A38351
phosphoprotein phosphatase 2-alpha regulatory chain - human
N;Alternate names: phosphoprotein phosphatase 2A-alpha 55K regulatory chain B
C;Species: Homo sapiens (man)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 05-Nov-1999
C;Accession: A38351
R:Mayer, R.E.; Hendrix, P.; Cron, P.; Matthies, R.; Stone, S.R.; Goris, J.; Merlevede
Biochemistry 30, 3589-3597, 1991
A;Title: Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: eviden
A;Reference number: A38351; MUID:9119806
A;Accession: A38351
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-447 <MAY>
A;Cross-references: GB:M64929; GB:J05328; NID:g190421; PIDN:AAA36490.1; PID:g190422
C;Genetics:
A;Gene: GDB:PPP2R2A
A;Cross-references: GDB:136800

```
            Query Match          41.3%;    Score 43;  DB  2;  Length 447;
            Best Local Similarity 50.0%;   Pred. No. 27;
            Matches    7;  Conservative  5;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1  SGSLSTFFRLFNRS 14
            :||.: |||:|:|:
Db    359  TGSYNNFFRMFDRN 372
```

Search completed: November   5, 2002, 10:56:44
Job time : 12.1493 secs

```
                GenCore version 5.1.3
        Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     November  5,  2002, 10:55:03 ; Search time 5.01493 Seconds
                                           (without alignments)
                                           162.138 Million cell updates/sec

Title:           US-09-833-017B-4
Perfect score:   104
Sequence:        1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:      105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	43.3	135	1	WN7C_XENLA	P31290 xenopus lae
2	44.5	42.8	689	1	GALX_PACTA	P40801 pachysolen
3	44	42.3	490	1	CPCR_MESAU	P33264 mesocricetu
4	44	42.3	934	1	ORP6_HUMAN	Q9bzf3 homo sapien
5	44	42.3	970	1	Y087_BUCAI	P57189 buchnera ap
6	44	41.3	230	1	FILL_YEAST	P38771 saccharomyc
7	43	41.3	413	1	?ABB_RABIT	Q00006 o serine/th
8	43	41.3	426	1	2ABA_PIG	Q29090 s serine/th
9	43	41.3	443	1	2ABB_HUMAN	Q00005 homo sapien
10	43	41.3	443	1	2ABB_PIG	P54614 sus scrofa
11	43	41.3	443	1	2ABB_RAT	P36877 r serine/th
12	43	41.3	447	1	2ABA_HUMAN	Q00007 h serine/th
13	43	41.3	447	1	2ABA_RAT	P36876 r serine/th
14	43	41.3	453	1	2ABD_RAT	P56932 r serine/th
15	43	41.3	479	1	5H2B_RAT	P30994 rattus norv
16	43	41.3	504	1	5H2B_MOUSE	Q02152 mus musculu
17	43	41.3	835	1	VP3_ROTS1	P15736 simian 11 r
18	43	41.3	1123	1	VG39_HSVI1	Q00143 ictalurid h
19	42	40.4	113	1	RAMA_KLEPN	Q48413 klebsiella
20	42	40.4	490	1	CPCP_MESAU	Q08078 mesocricetu
21	42	40.4	490	1	CPCQ_MESAU	P33263 mesocricetu
22	42	40.4	499	1	2ABA_DROME	P36872 drosophila
23	42	40.4	709	1	FTSH_MYCPN	P75120 mycoplasma
24	41.5	39.9	1242	1	MSH6_YEAST	Q03834 saccharomyc
25	41	39.4	113	1	RAMA_ENTCL	P55922 enterobacte
26	41	39.4	189	1	APOD_MOUSE	P51910 mus musculu
27	41	39.4	265	1	ISTB_PSEAE	P15026 pseudomonas
28	41	39.4	437	1	ADFP_HUMAN	Q99541 homo sapien
29	41	39.4	540	1	GUX1_ASPAC	O59843 aspergillus
30	41	39.4	2436	1	ABC2_HUMAN	Q9bzc7 homo sapien
31	40	38.5	96	1	5H2B_CAVPO	P97267 cavia porce
32	40	38.5	267	1	YO84_DEIRA	Q9rrk5 deinococcus
33	40	38.5	273	1	SCF_HUMAN	P21583 homo sapien
34	40	38.5	274	1	SCF_CANFA	Q06220 canis faml
35	40	38.5	447	1	2ABG_HUMAN	Q9y2t4 h serine/th
36	40	38.5	447	1	2ABG_RABIT	P50410 o serine/th
37	40	38.5	447	1	2ABG_RAT	P97888 r serine/th
38	40	38.5	481	1	5H2B_HUMAN	P41595 homo sapien
39	40	38.5	699	1	YQOA_CAEEL	Q09299 caenorhabdi
40	40	38.5	1222	1	YMP3_CAEEL	Q10947 caenorhabdi
41	40	38.5	1950	1	UBR1_YEAST	P19812 saccharomyc
42	39.5	38.0	521	1	VENV_DHV1	P27427 dhori virus
43	39.5	38.0	748	1	TAP1_HUMAN	Q03518 homo sapien
44	39	37.5	149	1	TSJT_TOBAC	P24805 nicotiana t
45	39	37.5	175	1	MAT1_YEAST	P01365 saccharomyc

```
                              ALIGNMENTS

RESULT 1
WN7C_XENLA
ID   WN7C_XENLA     STANDARD;      PRT;    135 AA.
AC   P31290;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   WNT-7C protein (XWNT-7C) (Fragment).
GN   WNT-7C.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9302368; PubMed=1408135;
RA   Wolda S.L., Moon R.T.;
RT   "Cloning and developmental expression in Xenopus laevis of seven
RT   additional members of the Wnt family.";
RL   Oncogene 7:1941-1947(1992).
CC   -!- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN
CC       TRANSMEMBRANE RECEPTORS. PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A
CC       SIGNALING MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE
CC       REGIONS OF TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL
CC       DIAMETERS.
CC   -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the
CC       extracellular matrix.
CC   -!- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L07535; AAA49988.1; .
DR   InterPro; IPR000970; Wnt1.
DR   Pfam; PF00110; wnt; 1.
DR   SMART; SM00097; WNT1; 1.
DR   PROSITE; PS00246; WNT1; 1.
KW   Developmental protein; Glycoprotein.
FT   NON_TER        1
FT   CARBOHYD      62     62       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      85     85       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      98     98       N-LINKED (GLCNAC...) (POTENTIAL).
FT   NON_TER      135    135
SQ   SEQUENCE    135 AA;  15344 MW;  8456A5DA03A39B4C CRC64;

Query Match             43.3%;  Score 45;  DB 1;  Length 135;
Best Local Similarity   50.0%;  Pred. No. 1.5;
Matches    9;  Conservative   4;  Mismatches   5;  Indels   0;  Gaps   0;

Qy          1 SGSLSTFFRLFNRSFTQA 18
```

```
Db    87  TGSIGTYGRFCNRTSTQA  104
          :||: |: | ||: |||
```

RESULT 2
GALX_PACTA
```
ID   GALX_PACTA    STANDARD;    PRT;    689 AA.
AC   P40801;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GAL10 bifunctional protein [Includes: UDP-glucose 4-epimerase
DE   (EC 5.1.3.2) (Galactowaldenase); Aldose 1-epimerase (EC 5.1.3.3)
DE   (Mutarotase)].
GN   GAL10.
OS   Pachysolen tannophilus (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pachysolen.
OX   NCBI_TaxID=4918;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-2460 / ATCC 32691;
RX   MEDLINE=94171069; PubMed=8125329;
RA   Skrzypek M., Maleszka R.;
RT   "A gene homologous to that encoding UDP galactose-4-epimerase is
RT   inducible by xylose in the yeast Pachysolen tannophilus.";
RL   Gene 140:127-129(1994).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.
CC   -!- CATALYTIC ACTIVITY: ALPHA-D-GLUCOSE = BETA-D-GLUCOSE (ALSO
CC       ACTS ON L-ARABINOSE, D-XYLOSE, D-GALACTOSE, MALTOSE AND LACTOSE).
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: GALACTOSE METABOLISM.
CC   -!- INDUCTION: BY XYLOSE.
CC   -!- SIMILARITY: IN THE N-TERMINAL PART WITH OTHER GALACTOWALDENASES,
CC       AND IN THE C-TERMINAL PART WITH OTHER MUTAROTASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X68593; CAA48580.1; -.
DR   PIR; S29621; S29621.
DR   HSSP; P09147; 1XEL.
DR   InterPro; IPR001823; Aldose_1_epimerase.
DR   InterPro; IPR001509; Epimerase.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF01263; Aldose_epim; 1.
DR   Pfam; PF01370; Epimerase; 1.
DR   ProDom; PD006407; Aldose_1_epimerase; 1.
DR   PROSITE; PS00545; ALDOSE_1_EPIMERASE; 1.
KW   Multifunctional enzyme; Isomerase; NAD; Galactose metabolism.
FT   DOMAIN       1    345       GALACTOWALDENASE.
FT   DOMAIN     346    689       MUTAROTASE.
FT   NP_BIND      3     34       NAD (POTENTIAL).
FT   ACT_SITE   534    534       MUTAROTASE (POTENTIAL).
SQ   SEQUENCE   689 AA;  76670 MW;  80C8A08262D3ED36 CRC64;

Query Match             42.8%;  Score 44.5;  DB 1;  Length 689;
Best Local Similarity   47.6%;  Pred. No. 10;
Matches   10;  Conservative   6;  Mismatches   4;  Indels   1;  Gaps   1;

Qy    1  SGSLSTFFRLFNRSFTQALGK  21
         :|: || :||: :| :|:||
Db  270  TGNGSTVFEVFN-AFCEAVGK  289
```

RESULT 3
CPCR_MESAU
```
ID   CPCR_MESAU    STANDARD;    PRT;    490 AA.
AC   P33264;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Cytochrome P450 2C27 (EC 1.14.14.1) (CYPIIC27) (P450 HSM3).
GN   CYP2C27.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94158799; PubMed=8114672;
RA   Sakuma T., Masaki K., Itoh S., Yokoi T., Kamataki T.;
RT   "Sex-related differences in the expression of cytochrome P450 in
RT   hamsters: cDNA cloning and examination of the expression of three
RT   distinct CYP2C cDNAs.";
RL   Mol. Pharmacol. 45:228-236(1994).
CC   -!- FUNCTION: CATALYZES THE HYDROXYLATION OF TOLBUTAMIDE AND THE N-
CC       DEMETHYLATION OF AMINOPYRINE AND BENZPHETAMINE.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED BY MALES IN KIDNEYS.
CC   -!- EXPRESSED PREDOMINANTLY BY MALES IN LIVERS.
CC   -!- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER
CC       TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,
CC       AND CARCINOGENS.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D11436; BAA02002.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   BINDING    435    435       HEME (BY SIMILARITY).
SQ   SEQUENCE   490 AA;  55767 MW;  450A208070D6DD0A CRC64;

Query Match             42.3%;  Score 44;  DB 1;  Length 490;
Best Local Similarity   42.9%;  Pred. No. 8.5;
Matches    9;  Conservative   5;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    1  SGSLSTFFRLFNRSFTQALGK  21
         :| ||| | :|| || |:
Db   51  SGSLTNFSKVYGPVFTLYLGR  71
```

RESULT 4
ORP6_HUMAN
```
ID   ORP6_HUMAN    STANDARD;    PRT;    934 AA.
AC   Q9BZF3; Q96SR1;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Oxysterol binding protein-related protein 6 (OSBP-related protein 6)
DE   (ORP-6).
GN   OSBPL6 OR ORP6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21776257; PubMed-11483621;
RA   Lehto M., Laitinen S., Chinetti G., Johansson M., Ehnholm C.,
RA   Staels B., Ikonen E., Olkkonen V.M.;
RT   "The OSBP-related protein family in humans.";
RL   J. Lipid Res. 42:1203-1213(2001).
RN   [2]
RP   SEQUENCE OF 216-934 FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE OSBP FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
```
```
DR   EMBL; AF323728; AAG53409.1; -.
DR   EMBL; AK027600; BAB55223.1; -.
DR   InterPro; IPR000648; Oxysterol_BP.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF01237; Oxysterol_BP; 1.
DR   Pfam; PF00169; PH; 1.
DR   SMART; SM00233; PH; 1.
DR   PROSITE; PS01013; OSBP; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Lipid transport; Transport.
FT   DOMAIN       86    181       PH.
FT   CONFLICT    319    319       S -> G (IN REF. 2).
FT   CONFLICT    714    714       N -> S (IN REF. 2).
SQ   SEQUENCE   934 AA;  106305 MW;  D20F90EA34C81497 CRC64;

Query Match             42.38%;  Score 44;  DB 1;  Length 934;
Best Local Similarity   42.98%;  Pred. No. 17;
Matches    9;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1  SGSLSTFFRLFNRSFTQALGK 21
              || ||||:|| ::|:| ||:
Db    634  SGYCSTYFRAGSKPFNPVLGE 654
```

RESULT 5
Y087_BUCAI
```
ID   Y087_BUCAI     STANDARD;      PRT;   970 AA.
AC   P57189;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein BU087.
GN   BU087.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOKYO 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- SUBUNIT: SOME, TO E.COLI YTFN.
```
```
DR   EMBL; AP001118; BAB12807.1; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     12     32       POTENTIAL.
SQ   SEQUENCE   970 AA;  114477 MW;  16B7BADB129F422E CRC64;

Query Match             42.3%;  Score 44;  DB 1;  Length 970;
Best Local Similarity   80.0%;  Pred. No. 17;
Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6  TFFRLFNRSF 15
              :|| ||||||
Db    703  SFFNLFNRSF 712
```

RESULT 6
FIL1_YEAST
```
ID   FIL1_YEAST     STANDARD;      PRT;   230 AA.
AC   P38771;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   FIL1 protein, mitochondrial precursor.
GN   FIL1 OR KIM4 OR YHR038W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT8-1;
RX   MEDLINE=9817448; PubMed=9746366;
RA   Kanai T., Takeshita S., Atomi H., Umemura K., Ueda M., Tanaka A.;
RT   "A regulatory factor, Fil1p, involved in derepression of the
RT   isocitrate lyase gene in Saccharomyces cerevisiae -- a possible
RT   mitochondrial protein necessary for protein synthesis in
RT   mitochondria.";
RL   Eur. J. Biochem. 256:212-220(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [3]
RP   SIMILARITY TO RRF.
RX   MEDLINE=96155066; PubMed=8563640;
RA   Ouzounis C., Bork P., Casari G., Sander C.;
RT   "New protein functions in yeast chromosome VIII.";
RL   Protein Sci. 4:2424-2428(1995).
CC   -!- FUNCTION: INVOLVED IN GLUCOSE REPRESSION/DEREPRESSION. MAY BE
CC       NECESSARY FOR PROTEIN SYNTHESIS IN MITOCHONDRIA. MAY FUNCTION AS A
CC       RIBOSOME RECYCLING FACTOR IN MITOCHONDRIA.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE RRF FAMILY.
```

CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AB016033; BAA31687.1; -.
DR   EMBL; U00062; AAB68906.1; -.
DR   PIR; S46737; S46737.
DR   SGD; S0001080; FIL1.
DR   InterPro; IPR002661; RRF.
DR   Pfam; PF01765; RRF; 1.
KW   Protein biosynthesis; Transit peptide; Mitochondrion.
FT   TRANSIT   1   24   MITOCHONDRION (POTENTIAL).
FT   CHAIN   25   230   FI11 PROTEIN.
SQ   SEQUENCE   230 AA; 26406 MW; 9CBCD8C5F86F3008 CRC64;

Query Match   41.3%; Score 43; DB 1; Length 230;
Best Local Similarity   80.0%; Pred. No. 5.6;
  Matches   8; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   9 RLFNRSFTQA 18
           ||||||:||:|
Db   17 RLFNRSFSQS 26

RESULT 7
2ABB_RABIT
ID   2ABB_RABIT   STANDARD;   PRT;   413 AA.
AC   Q00006;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE   beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B, B55-
DE   beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,
DE   R2-beta isoform) (Fragment).
GN   PPP2R2B.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=91198016; PubMed=1849734;
RA   Mayer R.E., Hendrix P., Cron P., Matthies R., Stone S.R.,
RA   Goris J., Merlevede W., Hofsteenge J., Hemmings B.A.;
RT   "Structure of the 55-kDa regulatory subunit of protein phosphatase
RT   2A: evidence for a neuronal-specific isoform.";
RL   Biochemistry 30:3589-3597(1991).
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC   SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC   LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC   COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC   COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC   CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATE
CC   WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC   WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC   B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
CC   FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC   AND CELL SIGNALING MOLECULES.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC   FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

---

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; M64931; AAA31458.1; -.
DR   InterPro; IPR000009; PP2A_PR55.
DR   Pfam; PF00400; WD40; 5.
DR   SMART; SM00320; WD40; 2.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.
KW   Multigene family.
FT   NON_TER   1   1
SQ   SEQUENCE   413 AA; 48243 MW; 49237B7817EB8FE2 CRC64;

Query Match   41.3%; Score 43; DB 1; Length 413;
Best Local Similarity   50.0%; Pred. No. 10;
  Matches   7; Conservative   5; Mismatches   2; Indels   0; Gaps   0;

Qy   1 SGSLSTFFRLFNRS 14
           :|| . ||||:|:|:
Db   325 TGSYNNFFRMFDRN 338

RESULT 8
2ABA_PIG
ID   2ABA_PIG   STANDARD;   PRT;   426 AA.
AC   Q29090;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE   alpha isoform (PP2A, subunit B, B-alpha isoform) (PP2A, subunit B,
DE   B55-alpha isoform) (PP2A, subunit B, PR55-alpha isoform) (PP2A,
DE   subunit B, R2-alpha isoform) (Fragment).
GN   PPP2R2A.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mayer-Jackel R.E.;
RL   Thesis (1992), Friedrich Miescher Institut / Basel, Switzerland.
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC   SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC   LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC   COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC   COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC   CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC   WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC   WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC   B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
CC   FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC   AND CELL SIGNALING MOLECULES.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC   FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z34932; CAA84404.1; -.
DR   InterPro; IPR000009; PP2A_PR55.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 5.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.

```
KW   Multigene family.
FT   NON_TER      1              1
SQ   SEQUENCE   426 AA;  49613 MW;  3AAD7EB338B03534 CRC64;

Query Match             41.3%;  Score 43;  DB 1;  Length 426;
Best Local Similarity   50.0%;  Pred. No. 11;
Matches    7;  Conservative   5;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1  SGSLSTFFRLFNRS 14
         :|| : |||:|:|:
Db  338  TGSYNNFFRMFDRN 351

RESULT 9
2ABB_HUMAN
ID   2ABB_HUMAN     STANDARD;      PRT;    443 AA.
AC   Q00005;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE   beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B, B55-
DE   beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,
DE   R2-beta isoform).
GN   PPP2R2B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=91198016; PubMed=1849734;
RA   Mayer R.E., Hendrix P., Cron P., Matthies R., Stone S.R.,
RA   Goris J., Merlevede W., Hofsteenge J., Hemmings B.A.;
RT   "Structure of the 55-kDa regulatory subunit of protein phosphatase
RT   2A: evidence for a neuronal-specific isoform.";
RL   Biochemistry 30:3589-3597(1991).
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC       SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC       LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC       COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC       COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC       CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC       WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC       WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC       B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC       FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC       AND CELL SIGNALING MOLECULES.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC       FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M64930; AAA36493.1; -.
DR   PIR; B38351; B38351.
DR   MIM; 604325;
DR   InterPro; IPR000009; PP2A_PR55.
DR   Pfam; PF00400; WD40; 6.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.
KW   Multigene family.
SQ   SEQUENCE   443 AA;  51710 MW;  C383C834B2852B8F CRC64;
```

```
Query Match             41.3%;  Score 43;  DB 1;  Length 443;
Best Local Similarity   50.0%;  Pred. No. 11;
Matches    7;  Conservative   5;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1  SGSLSTFFRLFNRS 14
         :|| : |||:|:|:
Db  355  TGSYNNFFRMFDRN 368

RESULT 10
2ABB_PIG
ID   2ABB_PIG       STANDARD;      PRT;    443 AA.
AC   P54614;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE   beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B, B55-
DE   beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,
DE   R2-beta isoform).
GN   PPP2R2B.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mayer-Jaekel R.E.;
RL   Thesis (1992), Friedrich Miescher Institut / Basel, Switzerland.
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC       SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC       LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC       COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC       COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC       CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC       WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC       WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC       B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC       FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC       AND CELL SIGNALING MOLECULES.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC       FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; Z34933; CAA84405.1; -.
DR   InterPro; IPR000009; PP2A_PR55.
DR   Pfam; PF00400; WD40; 5.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 4.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.
KW   Multigene family.
SQ   SEQUENCE   443 AA;  51459 MW;  F8562FC696719F41 CRC64;
```

```
RESULT 11
2ABB_RAT
ID   2ABB_RAT        STANDARD;      PRT;    443 AA.
AC   P36877;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE   beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B, B55-
DE   beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,
DE   R2-beta isoform) (PP2A, subunit B, BRB isoform).
GN   PPP2R2B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=95331316; PubMed=7607250;
RA   Akiyama N., Shima H., Hatano Y., Osawa Y., Sugimura T., Nagao M.;
RT   "cDNA cloning of BR gamma, a novel brain-specific isoform of the B
RT   regulatory subunit of type-2A protein phosphatase.";
RL   Eur. J. Biochem. 230:766-772(1995).
RN   [2]
RP   SEQUENCE OF 8-177 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=93779982; PubMed=8389301;
RA   Hatano Y., Shima H., Haneji T., Miura A.B., Sugimura T., Nagao M.;
RT   "Expression of PP2A B regulatory subunit beta isotype in rat testis.";
RL   FEBS Lett. 324:71-75(1993).
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC       SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC       LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC       COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC       COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC       CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC       WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC       WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC       B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC       FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC       AND CELL SIGNALING MOLECULES.
CC   -!- TISSUE SPECIFICITY: BRAIN AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D38260; BAA07412.1; -.
DR   EMBL; D14421; BAA03313.1; -.
DR   PIR; S33257; S33257.
DR   InterPro; IPR000009; PP2A_PR55.
DR   Pfam; PF00400; WD40; 6.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.
KW   Multigene family.
SQ   SEQUENCE  443 AA;  51668 MW;  0089EF6E8ED53082 CRC64;

Query Match             41.3%;  Score 43;  DB 1;  Length 443;
Best Local Similarity   50.0%;  Pred. No. 11;
Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SGSLSTFRLFNRS 14
           :|| : ||:|:|:
```

```
Db    355 TGSYNNFFRMFDRN 368

RESULT 12
2ABA_HUMAN
ID   2ABA_HUMAN      STANDARD;      PRT;    447 AA.
AC   Q00007; P50409;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 55 KDA regulatory subunit B,
DE   alpha isoform (PP2A, subunit B, B-alpha isoform) (PP2A, subunit B,
DE   B55-alpha isoform) (PP2A, subunit B, PR55-alpha isoform) (PP2A,
DE   subunit B, R2-alpha isoform)
GN   PPP2R2A.
OS   Homo sapiens (Human), and
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606, 9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human; TISSUE=Lung fibroblast;
RX   MEDLINE=91198016; PubMed=1849734;
RA   Mayer R.E., Hendrix P., Cron P., Matthies R., Stone S.R., Goris J.,
RA   Merlevede W., Hofstenge J., Hemmings B.A.;
RT   "Structure of the 55-kDa regulatory subunit of protein phosphatase
RT   2A: evidence for a neuronal-specific isoform.";
RL   Biochemistry 30:3589-3597(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Rabbit; STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;
RA   Depaoli-Roach A.A.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC       SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC       LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC       COMPARTMENT.
CC   -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC       COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC       CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC       WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC       WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC       B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC       FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC       AND CELL SIGNALING MOLECULES.
CC   -!- TISSUE SPECIFICITY: IN ALL TISSUES EXAMINED.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64929; AAA36490.1; -.
DR   EMBL; U09356; AAA18497.1; -.
DR   PIR; A38351; A38351.
DR   MIM; 604941; -.
DR   InterPro; IPR000009; PP2A_PR55.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 6.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS01024; PR55_1; 1.
DR   PROSITE; PS01025; PR55_2; 1.
KW   Multigene family.
SQ   SEQUENCE  447 AA;  51692 MW;  F4D407FF7ADA4ED6 CRC64;
```

Query Match   41.3%;   Score 43; DB 1; Length 447;
Best Local Similarity 50.0%;   Pred. No. 11;
Matches   7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy   1 SGSLSTFFRLFNRS 14
          :|| . ||:|:|:
Db 359 TGSYNNFFRMFDRN 372

RESULT 13
2ABA_RAT
ID 2ABA_RAT   STANDARD;   PRT;   447 AA.
AC P36876; P36878; O35512;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE alpha isoform (PP2A, subunit B, B-alpha isoform) (PP2A, subunit B,
DE B55-alpha isoform) (PP2A, subunit B, PR55-alpha isoform) (PP2A,
DE subunit B, R2-alpha isoform) (PP2A, subunit B, BRA isoform).
GN PP2R2A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114192; PubMed=1370560;
RA Pallas D.C., Weller W., Jaspers S., Miller T.B. Jr., Lane W.S.,
RA Roberts T.M.;
RT "The third subunit of protein phosphatase 2A (PP2A), a 55-kilodalton
RT protein which is apparently substituted for by T antigens in
RT complexes with the 36- and 63-kilodalton PP2A subunits, bears little
RT resemblance to T antigens.";
RL J. Virol. 66:886-893(1992).
RN [2]
RP SEQUENCE OF 80-272 FROM N.A.
RC STRAIN=FISCHER 344;
RX MEDLINE=93279382; PubMed=8389301;
RA Hatano Y., Shima H., Haneji T., Miura A.B., Sugimura T., Nagao M.;
RT "Expression of PP2A B regulatory subunit beta isotype in rat
RT testis.";
RL FEBS Lett. 324:71-75(1993).
CC -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC   SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC   LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC   COMPARTMENT.
CC -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC   COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC   CONSTANT REGULATORY SUBUNIT (PR65 or SUBUNIT A), THAT ASSOCIATES
CC   WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC   WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC   B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC   FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC   AND CELL SIGNALING MOLECULES.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC   FAMILY.
DR EMBL; M83298; AAA41910.1; -.
DR EMBL; M83297; AAA41909.1; -.
DR EMBL; D14419; BAA21904.1; -.
DR PIR; A41805; A41805.
DR InterPro; IPR000009; PP2A_PR55.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00600; PP2APR55.
DR SMART; SM00320; WD40; 3.
DR PROSITE; PS01024; PR55_1; 1.
DR PROSITE; PS01025; PR55_2; 1.
KW Multigene family.
FT VARIANT   60   60   E -> ESFKVHAALREASNLSMQ.
FT CONFLICT   105   105   K -> E (IN REF. 1; AAA41909).
FT CONFLICT   105   105   K -> R (IN REF. 2).
FT CONFLICT   213   213   N -> S (IN REF. 2).
FT CONFLICT   222   222   M -> V (IN REF. 2).
SQ SEQUENCE   447 AA; 51678 MW; 180AC837D9DA4ECE CRC64;

Query Match   41.3%;   Score 43; DB 1; Length 447;
Best Local Similarity 50.0%;   Pred. No. 11;
Matches   7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy   1 SGSLSTFFRLFNRS 14
          :|| . ||:|:|:
Db 359 TGSYNNFFRMFDRN 372

RESULT 14
2ABD_RAT
ID 2ABD_RAT   STANDARD;   PRT;   453 AA.
AC P56932;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE delta isoform (PP2A, subunit B, B-delta isoform) (PP2A, subunit B,
DE B55-delta isoform) (PP2A, subunit B, PR55-delta isoform) (PP2A,
DE subunit B, R2-delta isoform).
GN PPP2R2D.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=20026081; PubMed=10556517;
RA Strack S., Chang D., Zaucha J.A., Colbran R.J., Wadzinski B.E.;
RT "Cloning and characterization of B delta, a novel regulatory subunit
RT of protein phosphatase 2A.";
RL FEBS Lett. 460:462-466(1999).
CC -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC   SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC   LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC   COMPARTMENT.
CC -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC   COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC   CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES
CC   WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC   WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC   B (THE R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 AND R5/B'/B56
CC   FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC   AND CELL SIGNALING MOLECULES.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGH LEVELS IN BRAIN,
CC   HEART, PLACENTA, SKELETAL MUSCLE, TESTIS, THYMUS AND SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC   FAMILY.
DR EMBL; AF180350; AAF08536.1; -.

```
DR   InterPro; IPR000009; PP2A_PR55.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 6.
DR   PRINTS; PR00600; PP2APR55.
DR   SMART; SM00320; WD40; 2.
KW   Multigene family.
FT   DOMAIN        3        8       POLY-GLY.
SQ   SEQUENCE    453 AA;    51982 MW;  733E80A93A5BC2BB CRC64;

Query Match             41.3%;  Score 43;  DB 1;  Length 453;
Best Local Similarity   50.0%;  Pred. No. 11;
      Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SGSLSTFFRLFNRS 14
              :||:.:||:|::|:|
Db        365 TGSYNNFFRMFDRN 378


RESULT 15
5H2B_RAT
ID   5H2B_RAT        STANDARD;        PRT;       479 AA.
AC   P30994;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   5-hydroxytryptamine 2B receptor (5-HT-2B) (Serotonin receptor)
DE   (5-HT-2F) (Stomach fundus serotonin receptor).
GN   HTR2B OR SRL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=9371456; PubMed=1505525;
RA   Foguet M., Hoyer D., Pardo L.A., Parekh A., Kluxen F.-W.,
RA   Kalkman M.O., Stuehmer W., Luebbert H.;
RT   "Cloning and functional characterization of the rat stomach fundus
RT   serotonin receptor.";
RL   EMBO J. 11:3481-3487(1992).
CC   -!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR
CC       5-HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS
CC       AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THIS RECEPTOR
CC       MEDIATES ITS ACTION BY ASSOCIATION WITH G PROTEINS THAT ACTIVATE
CC       A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: STOMACH FUNDUS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       STRONGEST TO THE OTHER 5HT-2 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X66842; CAA47318.1; -.
DR   PIR; S23562; S23562.
DR   GCRDb; GCR_0434; -.
DR   InterPro; IPR000276; GPCR_Rhodopsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Lipoprotein; Palmitate.
FT   DOMAIN        1       55       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     56       78        1 (POTENTIAL).
FT   DOMAIN       79       90       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     91      112        2 (POTENTIAL).
```

```
FT   DOMAIN      113      128       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    129      150        3 (POTENTIAL).
FT   DOMAIN      151      170       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    171      191        4 (POTENTIAL).
FT   DOMAIN      192      215       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    216      238        5 (POTENTIAL).
FT   DOMAIN      239      323       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    324      344        6 (POTENTIAL).
FT   DOMAIN      345      359       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    360      382        7 (POTENTIAL).
FT   DOMAIN      383      479       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    127      206       BY SIMILARITY.
FT   CARBOHYD    203      203       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    353      353       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       396      396       PALMITATE (POTENTIAL).
SQ   SEQUENCE    479 AA;    53651 MW;  17FFC73213B42038 CRC64;

Query Match             41.3%;  Score 43;  DB 1;  Length 479;
Best Local Similarity   38.1%;  Pred. No. 12;
      Matches    8;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 SGSLSTFFRLFNRSFTQALGK 21
              || :|||::|::| |:|
Db        372 SGVNPLIYTLFNKTFREAFGR 392


Search completed: November  5, 2002, 10:57:55
Job time : 8.01493 secs
```

```
                    GenCore version 5.1.3
          Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   November  5, 2002, 10:55:02 ; Search time 13.4776 Seconds
                                 (without alignments)
                                 269.550 Million cell updates/sec

Title:            US-09-833-017B-4
Perfect score:    104
Sequence:         1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:   562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_19:*
              1: sp_archea:*
              2: sp_bacteria:*
              3: sp_fungi:*
              4: sp_human:*
              5: sp_invertebrate:*
              6: sp_mammal:*
              7: sp_mhc:*
              8: sp_organelle:*
              9: sp_phage:*
             10: sp_plant:*
             11: sp_rodent:*
             12: sp_virus:*
             13: sp_vertebrate:*
             14: sp_unclassified:*
             15: sp_rvirus:*
             16: sp_bacteriap:*
             17: sp_archeap:*

                         SUMMARIES
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	104	100.0	46	2	Q99QI5	Q99qi5 streptococc
2	104	100.0	46	2	Q9APK7	Q9apk7 streptococc
3	86	82.7	43	2	Q9APK6	Q9apk6 streptococc
4	48	46.2	1083	16	Q55584	Q55584 synechocyst
5	45	43.3	390	11	Q9QZS6	Q9zs6 mus musculu
6	45	43.3	835	12	Q9QNB1	Q9qnb1 human rotav
7	45	43.3	850	16	Q97N40	Q97n40 streptococc
8	45	43.3	1528	12	Q991N1	Q991n1 little cher
9	45	43.3	1980	12	Q991N2	Q991n2 little cher
10	44	42.3	311	4	Q96RX7	Q96rx7 homo sapien
11	44	42.3	381	4	Q96QI5	Q96qi5 homo sapien
12	44	42.3	431	5	Q9N3H3	Q9n3h3 caenorhabdi
13	44	42.3	719	4	Q96SR1	Q96sr1 homo sapien
14	44	42.3	740	6	Q9N007	Q9n007 macaca fasc
15	44	42.3	799	11	Q9DBI0	Q9dbi0 mus musculu
16	44	42.3	844	6	Q9N006	Q9n006 macaca fasc
17	44	42.3	889	10	Q9SKE7	Q9ske7 arabidopsis
18	44	42.3	934	4	Q9BZF3	Q9bzf3 homo sapien
19	43	41.3	113	2	Q9K5A5	Q9k5a5 enterobacte
20	43	41.3	167	16	Q99WL1	Q99wl1 staphylococ
21	43	41.3	305	16	O31443	O31443 bacillus su
22	43	41.3	432	6	Q95LX5	Q95lx5 macaca fasc
23	43	41.3	445	11	Q9D3B7	Q9d3b7 mus musculu
24	43	41.3	446	11	Q9D6I1	Q9d6i1 mus musculu
25	43	41.3	453	11	Q925E7	Q925e7 mus musculu
26	43	41.3	457	11	Q9CW40	Q9cw40 mus musculu
27	43	41.3	468	13	P87345	P87345 xenopus lae
28	43	41.3	468	13	Q92141	Q92141 xenopus lae
29	43	41.3	479	11	Q9QW44	Q9qw44 rattus sp.
30	43	41.3	479	11	Q9QWS2	Q9qws2 mus musculu
31	42.5	40.9	748	4	Q16149	Q16149 homo sapien
32	42	40.4	139	17	O59457	O59457 pyrococcus
33	42	40.4	142	9	Q9MBS1	Q9mbs1 staphylococ
34	42	40.4	155	5	O16597	O16597 caenorhabdi
35	42	40.4	206	5	Q9JRA2	Q9jra2 neisseria m
36	42	40.4	225	5	Q9VH21	Q9vh21 drosophila
37	42	40.4	393	16	Q9KTX2	Q9ktx2 vibrio chol
38	42	40.4	477	4	Q9P1Y7	Q9p1y7 homo sapien
39	42	40.4	514	16	Q9CKM5	Q9ckm5 pasteurella
40	42	40.4	608	11	Q9QXK1	Q9qxk1 mus musculu
41	42	40.4	665	8	Q9G8V7	Q9g8v7 rhodomonas
42	42	40.4	851	16	Q9I442	Q9i442 pseudomonas
43	41	39.4	136	2	O82915	O82915 escherichia
44	41	39.4	158	17	O58392	O58392 pyrococcus
45	41	39.4	210	11	Q9CU40	Q9cu40 mus musculu

ALIGNMENTS

```
RESULT 1
Q99QI5
ID   Q99QI5;          PRELIMINARY;       PRT;   46 AA.
AC   Q99QI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN PRECURSOR.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB14, H7, LT11, NG8, AND UA159;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277152; AAK01542.1; -.
DR   EMBL; AF277153; AAK01543.1; -.
DR   EMBL; AF277155; AAK01545.1; -.
DR   EMBL; AF277156; AAK01546.1; -.
DR   EMBL; AF277157; AAK01547.1; -.
DR   InterPro; IPR004288; ComC.
DR   Pfam; PF03047; ComC; 1.
FT   CHAIN       26     46       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;   5211 MW;  38FA62B6F78FC3BF CRC64;

Query Match       100.0%;  Score 104;  DB 2;  Length 46;
Best Local Similarity   100.0%;  Pred. No. 1.7e-10;
Matches   21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 SGSLSTFRLFNRSFTQALGK 21
           |||||||||||||||||||||
Db    26 SGSLSTFRLFNRSFTQALGK 46
```

```
RESULT 2
Q9APK7
ID   Q9APK7        PRELIMINARY;        PRT;        46 AA.
AC   Q9APK7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BM71;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF271151; AAK01541.1; -.
FT   CHAIN         26     46       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;   5195 MW;  38E0B9D0D5B8FC3BF CRC64;

Query Match             100.0%;  Score 104; DB 2;  Length 46;
Best Local Similarity   100.0%;  Pred. No. 1.7e-10;
Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SGSLSTFFRLFNRSFTQALGK 21
              |||||||||||||||||||||
Db         26 SGSLSTFFRLFNRSFTQALGK 46


RESULT 3
Q9APK6
ID   Q9APK6        PRELIMINARY;        PRT;        43 AA.
AC   Q9APK6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH1005;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF271154; AAK01544.1; -.
FT   CHAIN         26     43       COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   43 AA;   4927 MW;  E6A78FC3BF6156C7 CRC64;

Query Match             82.7%;   Score 86; DB 2;  Length 43;
Best Local Similarity   94.4%;   Pred. No. 1.8e-07;
Matches    17;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SGSLSTFFRLFNRSFTQA 18
              ||:|||||||||||||||
Db         26 SGTLSTFFRLFNRSFTQA 43


RESULT 4
Q55584
ID   Q55584        PRELIMINARY;        PRT;       1083 AA.
AC   Q55584;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACRIFLAVIN RESISTANCE PROTEIN.
GN   ENVD OR SLR0369.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 6803;
RA   Tabata S.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D63999; BAA10089.1; -.
DR   InterPro; IPR001036; ACR_tran.
DR   Pfam; PF00873; ACR_tran; 1.
DR   PRINTS; PR00702; ACRIFLAVINRP.
KW   Complete Proteome.
SQ   SEQUENCE   1083 AA;  117561 MW;  4388B790D6BC177A CRC64;

Query Match             46.2%;   Score 48; DB 16;  Length 1083;
Best Local Similarity   64.3%;   Pred. No. 17;
Matches    9;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GSLSTFFRLFNRSF 15
              |:||:||||||:|:|
Db        514 GPLAWFNLFNRTF 527


RESULT 5
Q9QZS6
ID   Q9QZS6        PRELIMINARY;        PRT;       390 AA.
AC   Q9QZS6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D-GLYCOSAMINYL 3-O-SULFOTRANSFERASE-3B.
GN   HS3ST3B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=99449300; PubMed=10520990;
RA   Shukla D., Liu J., Blaiklock P., Shworak N.W., Bai X., Esko J.D.,
RA   Cohen G.H., Eisenberg R.J., Rosenberg R.D., Spear P.G.;
RT   "A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus
RT   1 entry.";
RL   Cell 99:13-22(1999).
DR   EMBL; AF168992; AAF04505.1; -.
```

```
DR   MGD; MGI:1333853; Hs3st3b.
DR   InterPro; IPR000863; Sulfotransferase.
DR   Pfam; PF00685; Sulfotransfer; 1.
KW   Transferase.
SQ   SEQUENCE   390 AA;  43326 MW;  ACD0D28D66B3DDE8 CRC64;

Query Match             43.3%;  Score 45;  DB 11;  Length 390;
Best Local Similarity   50.0%;
Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LSTFFRLFNRSFTQALGK 21
              |:|:|||||| |:
Db        368 LRDFYRPFNRKFYQMTGR 385

RESULT 6
Q9QNB1
ID   Q9QNB1      PRELIMINARY;     PRT;   835 AA.
AC   Q9QNB1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VP3.
GN   VP3.
OS   Human rotavirus (strain KU).
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=10952;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KU;
RA   Taniguchi K.;
RT   "Rotavirus.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB022767; BAA84964.1; -.
SQ   SEQUENCE   835 AA;  97857 MW;  EA5D8227C147E599 CRC64;

Query Match             43.3%;  Score 45;  DB 12;  Length 835;
Best Local Similarity   66.7%;  Pred. No. 42;
Matches    8;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    0;

Qy          1 SGGLSTFFRLFN 12
              || |||:|:|:|
Db        768 SGVLSTYFKLYN 779

RESULT 7
Q97N40
ID   Q97N40      PRELIMINARY;     PRT;   850 AA.
AC   Q97N40;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
GN   SP2231.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
```

```
DR   EMBL; AE007510; AAK76279.1; -.
DR   TIGR; SP2231; -.
KW   Complete proteome.
SQ   SEQUENCE   850 AA;  97303 MW;  1ADED613F06B5115 CRC64;

Query Match             43.3%;  Score 45;  DB 16;  Length 850;
Best Local Similarity   50.0%;  Pred. No. 43;
Matches    8;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          6 TFFRLFNRSFTQALGK 21
              ||:|| :|||:||:|
Db        733 TFYRLDTKTFTEAIQK 748

RESULT 8
Q991N1
ID   Q991N1      PRELIMINARY;     PRT;   1528 AA.
AC   Q991N1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HELICASE (FRAGMENT).
OS   Little cherry virus-2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=154339;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rott M.E., Jelkmann W.;
RT   "Identification of a second closterovirus associated with little
RT   cherry disease, little cherry virus-2.";
RL   Phytopathology 91:0-0(2001).
DR   EMBL; AF333237; AAK19543.1; -.
DR   InterPro; IPR000606; Viral_helicase1.
DR   InterPro; IPR002588; V_methyltransf.
DR   Pfam; PF01443; Viral_helicase1; 1.
DR   Pfam; PF01660; Vmethyltransf; 1.
FT   NON_TER       1
SQ   SEQUENCE   1528 AA;  170263 MW;  1AA54A7016AE27E2 CRC64;

Query Match             43.3%;  Score 45;  DB 12;  Length 1528;
Best Local Similarity   45.0%;  Pred. No. 80;
Matches    9;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GSLSTFFRLFNRSFTQALGK 21
              ||| |: ||| :|| ::|
Db        325 GSLVGMFHMVGRRFTNTIGK 344

RESULT 9
Q991N2
ID   Q991N2      PRELIMINARY;     PRT;   1980 AA.
AC   Q991N2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA DEPENDENT RNA POLYMERASE (FRAGMENT).
OS   Little cherry virus-2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=154339;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rott M.E., Jelkmann W.;
RT   "Identification of a second closterovirus associated with little
RT   cherry disease, little cherry virus-2.";
RL   Phytopathology 91:0-0(2001).
DR   EMBL; AF333237; AAK19544.1; -.
DR   InterPro; IPR001788; RNA_dep_RNApol2.
DR   InterPro; IPR000606; Viral_helicase1.
DR   InterPro; IPR002588; V_methyltransf.
DR   Pfam; PF00978; RNA_dep_RNApol2; 1.
DR   Pfam; PF01443; Viral_helicase1; 1.
```

```
DR   Pfam; PF01660; Vmethyltransf; 1.
FT   NON_TER          1            1
FT   NON_TER       1980         1980
SQ   SEQUENCE   1980 AA;  221780 MW;  10A6C535BBB6D611 CRC64;

Query Match             43.3%;  Score 45;  DB 12;  Length 1980;
Best Local Similarity   45.0%;  Pred. No. 1.1e+02;
Matches    9;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      2 GSLSTFFRLFNRSFTQALGK 21
          ||| .|: .|| ||| :||
Db    325 GSLVGMFHMVGRRFTNTIGK 344


RESULT 10
Q96RX7
ID   Q96RX7           PRELIMINARY;       PRT;   311 AA.
AC   Q96RX7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HEPARAN SULPHATE D-GLUCOSAMINYL 3-O-SULFOTRANSFERASE-3B
DE   LIKE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21096910; PubMed=11157797;
RA   Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA   Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA   Higgs D.R.;
RT   "Sequence, structure and pathology of the fully annotated terminal 2
RT   Mb of the short arm of human chromosome 16.";
RL   Hum. Mol. Genet. 10:339-352(2001).
DR   EMBL; AE006640; AAK61299.1; -.
KW   Transferase.
SQ   SEQUENCE   311 AA;  34694 MW;  6841E6151BA0DA6F CRC64;

Query Match             42.3%;  Score 44;  DB 4;  Length 311;
Best Local Similarity   50.0%;  Pred. No. 21;
Matches    9;  Conservative    2;  Mismatches    2;  Indels    7;  Gaps    0;

Qy      4 LSTFFRLFNRSFTQALGK 21
          | :|| |||| | :
Db    289 LQEFYRPFNRRFYQMTGQ 306


RESULT 11
Q96QI5
ID   Q96QI5           PRELIMINARY;       PRT;   381 AA.
AC   Q96QI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C439A6.1 (NOVEL PROTEIN SIMILAR TO HEPARAN SULFATE (GLUCOSAMINE)
DE   3-O-SULFOTRANSFERASES) (FRAGMENT).
GN   C439A6.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Thomas D.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031723; CAC42157.1; -.
FT   NON_TER          1
SQ   SEQUENCE   381 AA;  41205 MW;  3899BECFE0218285 CRC64;

Query Match             42.3%;  Score 44;  DB 4;  Length 381;
Best Local Similarity   50.0%;  Pred. No. 26;
```

```
Matches    9;  Conservative    2;  Mismatches    2;  Indels    7;  Gaps    0;

Qy      4 LSTFFRLFNRSFTQALGK 21
          | :|| |||| | :
Db    359 LQEFYRPFNRRFYQMTGQ 376


RESULT 12
Q9N3H3
ID   Q9N3H3           PRELIMINARY;       PRT;   431 AA.
AC   Q9N3H3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 48.3 KDA PROTEIN.
GN   Y53G8AL.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=9069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Latreille P.;
RT   "The sequence of C. elegans cosmid Y53G8AL.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC024807; AAF59528.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   431 AA;  48254 MW;  D5ADA2DDC3952A2C CRC64;

Query Match             42.3%;  Score 44;  DB 5;  Length 431;
Best Local Similarity   38.1%;  Pred. No. 30;
Matches    8;  Conservative    4;  Mismatches    4;  Indels    9;  Gaps    0;

Qy      1 SGSLSTFFRLFNRSFTQALGK 21
          :|| || ||| : |:
Db    375 AGGQQAFYRSFNRYFEEQYGE 395


RESULT 13
Q96SR1
ID   Q96SR1           PRELIMINARY;       PRT;   719 AA.
AC   Q96SR1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ14694 FIS, CLONE NT2RP2005407, WEAKLY SIMILAR TO
DE   OXYSTEROL-BINDING PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yanamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
```

```
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK027600; BAB55223.1; -.
SQ   SEQUENCE   719 AA;   81876 MW;   A2B1OD14D265A41E CRC64;

Query Match
Best Local Similarity   42.3%;      Score 44;  DB 4;  Length 719;
                        42.9%;      Pred. No. 53;  -
Matches  9;   Conservative  4;   Mismatches  4;   Indels  8;   Gaps  0;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
         || ||:|| :::|| :| ||:
Db  419 SGYCSTYFRAGSKPFNPVLGE 439

RESULT 14
Q9N007
ID   Q9N007;               PRELIMINARY;       PRT;    740 AA.
AC   Q9N007;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 84.2 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX.
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB046640; BAB03558.1; -.
DR   InterPro; IPR000648; Oxysterol_BP.
DR   Pfam; PF01237; Oxysterol_BP; 1.
DR   PROSITE; PS01013; OSBP; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   740 AA;   84223 MW;   6B8E850FD9E13C78 CRC64;

Query Match
Best Local Similarity   42.3%;      Score 44;  DB 6;  Length 740;
                        42.9%;      Pred. No. 54;  -
Matches  9;   Conservative  4;   Mismatches  8;   Indels  0;   Gaps  0;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
         || ||:|| :::|| :| ||:
Db  440 SGYCSTYFRAGSKPFNPVLGE 460

RESULT 15
Q9DBI0
ID   Q9DBI0;               PRELIMINARY;       PRT;    799 AA.
AC   Q9DBI0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   1300008A22RIK PROTEIN.
GN   1300008A22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LIVER;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR   EMBL; AK004939; BAB23684.1; -.
DR   HSSP; P00763; 1DPO.
DR   MGD; MGI:1919003; 1300008A22Rik.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00057; ldl_recept_a; 3.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00042; CUB; 1.
DR   SMART; SM00192; LDLa; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS50068; LDLRA_2; 3.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW   Glycoprotein; Hydrolase; Serine protease.
SQ   SEQUENCE   799 AA;   89557 MW;   1631A6464A4D5288 CRC64;

Query Match
Best Local Similarity   42.3%;      Score 44;  DB 11;  Length 799;
                        52.4%;      Pred. No. 59;  -
Matches  11;   Conservative  3;   Mismatches  3;   Indels  4;   Gaps  1;

Qy    1 SGSLSTFFRLFNRSFTQALGK 21
         |||  ||:|| ||: ||: | ||
Db   79 SGSL---RVLNRHFSQDLGR 95

Search completed: November  5, 2002, 10:57:35
Job time : 18.4776 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: November 5, 2002, 10:57:42 ; Search time 16.6119 seconds
(without alignments)
140.414 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 21
Sequence: 1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size :   0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	33.3	929	22	ABB70857	Drosophila melanog
2	6	28.6	27	20	AAY40080	Peptide sequence d
3	6	28.6	48	22	AAM23675	Human EST encoded
4	6	28.6	51	17	AAR94979	Calcium channel is
5	6	28.6	51	17	AAR94978	Calcium channel is
6	6	28.6	62	17	AAR97851	Rat brain calcium
7	6	28.6	66	17	AAR97852	Rat brain calcium
8	6	28.6	66	22	ABG04703	Novel human diagno
9	6	28.6	78	22	AAM89377	Human immune/haema
10	6	28.6	96	19	AAW59796	Nucleotide sequenc
11	6	28.6	122	20	AAY40074	Peptide sequence d

No.	Description	ID		Length	Score	
12	Propionibacterium	AAU45360	22	141	28.6	6
13	Human CCAAT/enhanc	AAE11946	22	147	28.6	6
14	Human novel secret	AAU16099	22	150	28.6	6
15	Arabidopsis thalia	AAG48562	21	152	28.6	6
16	Amino acid sequenc	AAW59797	19	180	28.6	6
17	Arabidopsis thalia	AAG19904	22	186	28.6	6
18	Human diagnostic a	AAU19473	22	194	28.6	6
19	Novel human diagno	ABG24830	20	212	28.6	6
20	Chlamydia trachoma	AAY37757	22	222	28.6	6
21	Novel human diagno	ABG12787	22	254	28.6	6
22	Novel human diagno	ABG15795	21	257	28.6	6
23	Arabidopsis thalia	AAG19903	13	270	28.6	6
24	Human calcium chan	AAR27641	22	279	28.6	6
25	Human nervous syst	ABB16367	22	279	28.6	6
26	Human novel secret	AAU16524	22	282	28.6	6
27	Chicken stem cell	AAU05262	22	282	28.6	6
28	Chicken SCF protei	AAB98364	22	282	28.6	6
29	Chicken SCF (stem	AAU02489	22	282	28.6	6
30	Chicken SCF (stem	AAU02775	22	282	28.6	6
31	Chicken SCF (stem	AAB73576	22	282	28.6	6
32	Chicken stem cell	AAB96949	22	286	28.6	6
33	Novel human secret	AAU32713	21	287	28.6	6
34	Arabidopsis thalia	AAG19902	22	296	28.6	6
35	Mouse CCAAT/enhanc	AAE11947	22	297	28.6	6
36	Rat CCAAT/enhancer	AAE11948	22	322	28.6	6
37	Human CCAAT/enhanc	AAE11945	22	329	28.6	6
38	Haemophilus influe	AAU35549	22	329	28.6	6
39	Novel human diagno	ABG11238	22	334	28.6	6
40	Propionibacterium	AAU67978	22	344	28.6	6
41	Novel human diagno	ABG03199	12	345	28.6	6
42	Nuclear factor C/E	AAR14408	22	345	28.6	6
43	Human CCAAT/enhanc	AAE11944	22	345	28.6	6
44	Human transcriptio	AAU09068	22	356	28.6	6
45	Human polypeptide	AAM39772	22	360	28.6	6
46	Drosophila melanog	ABB70995	20	359	28.6	6
47	Chlamydia pneumoni	AAY34924	21	360	28.6	6
48	Amino acid sequenc	AAY69364	22	360	28.6	6
49	Novel human diagno	ABG14697	22	360	28.6	6
50	Human polypeptide	AAM41558	22	372	28.6	6
51	Propionibacterium	AAU48589	22	398	28.6	6
52	Novel human diagno	ABG10841	22	412	28.6	6
53	Arabidopsis thalia	AAG48540	21	414	28.6	6
54	Arabidopsis thalia	AAG48539	21	446	28.6	6
55	Mature secreted hum	AAY30792	20	447	28.6	6
56	Arabidopsis thalia	AAG48538	21	465	28.6	6
57	Novel human diagno	ABG03158	20	466	28.6	6
58	Immature human HKN	AAY30793	20	477	28.6	6
59	Protein encoded by	AAY30786	22	477	28.6	6
60	Human protein sequ	AAB94726	22	478	28.6	6
61	Drosophila melanog	ABB58996	13	489	28.6	6
62	Human calcium chan	AAR27647	20	495	28.6	6
63	Protein encoded by	AAY30785	13	499	28.6	6
64	The ABF-B from A.	AAR27574	13	513	28.6	6
65	Human polypeptide	AAM1685	21	518	28.6	6
66	Human calcium chan	AAR27644	22	559	28.6	6
67	Human chaperone pr	AAY91945	22	559	28.6	6
68	Human protein sequ	AAB93115	21	561	28.6	6
69	E. coli cellular p	AAU34618	22	595	28.6	6
70	Salmonella typhi c	AAU38334	20	600	28.6	6
71	Zea mays protein f	AAG40930	21	635	28.6	6
72	Human secretory pr	AAW75855	19	635	28.6	6
73	Human secreted pro	AAY45160	21	635	28.6	6
74	Human membrane tra	AAY71060	13	684	28.6	6
75	Human calcium chan	AAR27645	22	698	28.6	6
76	Human GTPase assoc	AAY99666	15	703	28.6	6
77	Sequence of lag D	AAR49135	13	908	28.6	6
78	Human bone marrow	AAM00834	22	993	28.6	6
79	Human calcium chan	AAR27650	22	1087	28.6	6
80	Novel human diagno	ABG24996	22	1127	28.6	6
81	Human calcium chan	ABG22813	22	1217	28.6	6
82	Novel human diagno	AAU33177	22	1343	28.6	6
83	Novel human secret	AAM00947	22	1359	28.6	6
84	Human bone marrow	AAB41785	21		28.6	6

Index	Window	Score	Pos	Count	Accession	Description
85	6	28.6	1820	22	AAU27679	Human full-length
86	6	28.6	1824	21	AAY49431	Murine CACNAIF pro
87	6	28.6	1912	21	AAY49429	Human CACNAIF spli
88	6	28.6	1967	14	AAR33547	Sequence of the al
89	6	28.6	1968	19	AAW63139	Human calcium chan
90	6	28.6	1977	21	AAY49430	Human CACNAIF long
91	6	28.6	2138	16	AAR72607	Human neuronal cal
92	6	28.6	2138	21	AAB10593	Human calcium chan
93	6	28.6	2161	14	AAR33545	Sequence of the al
94	6	28.6	2161	16	AAR71001	Human neuronal cal
95	6	28.6	2161	16	AAR71002	Human neuronal cal
96	6	28.6	2161	19	AAW63137	Human calcium chan
97	6	28.6	2161	19	AAW63149	Human calcium chan
98	6	28.6	2161	21	AAB10568	Human calcium chan
99	6	28.6	2163	16	AAR71003	Human neuronal cal
100	6	28.6	2163	21	AAB10570	Human calcium chan
101	6	28.6	2279	22	AAM38733	Human polypeptide
102	6	28.6	2297	22	AAM40519	Human polypeptide
103	6	28.6	2322	18	AAW15566	Melanoma-associate
104	6	28.6	2344	22	AAU37120	Staphylococcus aur
105	5	23.8	5	14	AAR44624	Calcitonin fragmen
106	5	23.8	8	18	AAW13431	Melanoma-homing pe
107	5	23.8	8	19	AAW60292	Melanoma tumour ho
108	5	23.8	8	20	AAW93716	Mouse B16B15b mela
109	5	23.8	8	21	AAB21704	Murine melanoma ho
110	5	23.8	8	21	AAB23539	Angiogenic vascula
111	5	23.8	8	21	AAB17961	Integrin-binding p
112	5	23.8	8	22	AAE06282	Tumour homing pept
113	5	23.8	13	14	AAR40791	Sequence #4 from a
114	5	23.8	15	20	AAY22379	TPO receptor bindi
115	5	23.8	15	20	AAW92605	Mouse beta-actin a
116	5	23.8	15	20	AAW92534	Beta-actin referen
117	5	23.8	16	15	AAR47026	Plasminogen activa
118	5	23.8	19	19	AAW69830	Peptide fragment p
119	5	23.8	19	21	AAY85614	Cathepsin-D antige
120	5	23.8	21	18	AAW34053	Human MDR1-P glyco
121	5	23.8	23	16	AAR65099	Random biotinylati
122	5	23.8	23	19	AAW46702	Biotinylation pept
123	5	23.8	23	20	AAY29305	Biotinylation pept
124	5	23.8	23	20	AAW67910	Human secreted pro
125	5	23.8	23	21	AAY87142	Human secreted pro
126	5	23.8	23	21	AAY54984	Angiogenin inhibit
127	5	23.8	23	22	AAU04311	ATP-binding casset
128	5	23.8	23	22	AAE06119	Human gene 14 enco
129	5	23.8	23	22	AAE09937	Biotinylated pepti
130	5	23.8	25	22	ABG01293	Novel human diagno
131	5	23.8	27	20	AAW67973	Fragment of human
132	5	23.8	30	20	AAQ08982	Human polypeptide
133	5	23.8	31	7	AAP61348	Sequence of calcit
134	5	23.8	31	7	AAP61349	Sequence of calcit
135	5	23.8	31	22	AAM75470	Human bone marrow
136	5	23.8	32	17	AAW03618	Rat bradykinin GPR
137	5	23.8	33	22	AAG92939	C glutamicum prote
138	5	23.8	34	22	ABG08616	Novel human diagno
139	5	23.8	35	22	ABG15474	Novel human diagno
140	5	23.8	35	21	AAB18882	Peptide #5316 enco
141	5	23.8	36	21	AAB38977	Human secreted pep
142	5	23.8	36	22	ABG10490	Novel human diagno
143	5	23.8	39	18	AAW27950	Novel human diagno
144	5	23.8	40	9	AAP80076	Cytomegalovirus (C
145	5	23.8	40	21	AAG55145	Arabidopsis thalia
146	5	23.8	40	22	ABB22221	Protein #4220 enco
147	5	23.8	40	22	AAU01625	Human secreted pro
148	5	23.8	41	21	AAB50421	Human uncoupling p
149	5	23.8	41	21	AAB27164	HPIV2 partial prot
150	5	23.8	41	22	ABB31202	Peptide #3853 enco
151	5	23.8	41	22	ABB36400	Peptide #3906 enco
152	5	23.8	41	22	AAM57165	Human brain expres
153	5	23.8	41	22	AAM69564	Human bone marrow
154	5	23.8	41	22	AAM17385	Peptide #3819 enco
155	5	23.8	41	22	AAM29904	Peptide #3941 enco
156	5	23.8	41	22	AAM05074	Peptide #3756 enco
157	5	23.8	43	22	AAB82905	Human TNF receptor
158	5	23.8	44	22	ABG28609	Novel human diagno
159	5	23.8	46	20	AAY30849	Human secreted pro
160	5	23.8	46	22	AAM89492	Human immune/haema
161	5	23.8	46	22	AAG76210	Human colon cancer
162	5	23.8	47	22	ABB41431	Peptide #8937 enco
163	5	23.8	47	22	AAM62304	Human brain expres
164	5	23.8	47	22	AAM75107	Human bone marrow
165	5	23.8	47	22	AAM35223	Peptide #9260 enco
166	5	23.8	48	22	AAG76454	Human colon cancer
167	5	23.8	48	22	ABB42096	Peptide #9602 enco
168	5	23.8	49	22	AAM62976	Human brain expres
169	5	23.8	49	22	AAM75789	Human bone marrow
170	5	23.8	49	22	AAM35899	Peptide #9936 enco
171	5	23.8	50	22	AAU50306	Propionibacterium
172	5	23.8	51	17	AAR94977	Calcium channel is
173	5	23.8	52	21	AAG00521	Human secreted pro
174	5	23.8	52	22	ABB16203	Human nervous syst
175	5	23.8	52	22	AAM06584	Human foetal prote
176	5	23.8	52	22	AAM06729	Human foetal prote
177	5	23.8	54	21	AAB38489	Human secreted pro
178	5	23.8	54	22	ABB17231	Human nervous syst
179	5	23.8	55	22	ABB29477	Peptide #2128 enco
180	5	23.8	55	22	ABB31000	Peptide #3651 enco
181	5	23.8	55	22	ABB34655	Peptide #2161 enco
182	5	23.8	55	22	ABB36193	Peptide #3699 enco
183	5	23.8	55	22	ABB15421	Human nervous syst
184	5	23.8	55	22	ABB20062	Protein #2061 enco
185	5	23.8	55	22	AAM55445	Human brain expres
186	5	23.8	55	22	AAM56974	Human brain expres
187	5	23.8	55	22	AAM60182	Human brain expres
188	5	23.8	55	22	AAM67836	Human bone marrow
189	5	23.8	55	21	AAM69361	Human bone marrow
190	5	23.8	55	22	AAM72800	Human bone marrow
191	5	23.8	55	22	AAM15651	Peptide #2085 enco
192	5	23.8	55	22	AAM28152	Peptide #2189 enco
193	5	23.8	55	22	AAM33030	Peptide #7067 enco
194	5	23.8	55	22	AAM03390	Peptide #2072 enco
195	5	23.8	55	22	AAM04888	Peptide #3570 enco
196	5	23.8	57	22	AAM23726	Human EST encoded
197	5	23.8	57	22	AAM24492	Human EST encoded
198	5	23.8	57	21	AAG02747	Human secreted pro
199	5	23.8	60	22	AAU56956	Propionibacterium
200	5	23.8	60	21	ABG04669	Propionibacterium
201	5	23.8	61	22	ABG09797	Novel human diagno
202	5	23.8	61	22	AAM83024	Human immune/haema
203	5	23.8	62	21	AAM34466	Human secreted pro
204	5	23.8	62	22	AAG91645	C glutamicum prote
205	5	23.8	63	21	AAG187764	Zea mays protein f
206	5	23.8	63	21	AAG32669	Zea mays protein f
207	5	23.8	63	21	AAG48401	Arabidopsis thalia
208	5	23.8	63	22	AAM81044	Human haematologic
209	5	23.8	63	22	AAM82041	Human haematologic
210	5	23.8	63	22	AAM95167	Human reproductive
211	5	23.8	65	22	AAU58313	Propionibacterium
212	5	23.8	65	22	ABB39031	Peptide #6537 enco
213	5	23.8	65	22	ABB23975	Protein #5974 enco
214	5	23.8	65	22	AAM59665	Human brain expres
215	5	23.8	65	22	AAM72261	Human bone marrow
216	5	23.8	65	22	AAM19520	Peptide #5954 enco
217	5	23.8	65	22	AAM32520	Peptide #6557 enco
218	5	23.8	65	22	AAM41112	Human polypeptide
219	5	23.8	66	20	AAY60064	Human endometrium
220	5	23.8	66	22	ABB44134	Peptide #11640 enc
221	5	23.8	66	22	ABB23022	Protein #5021 enco
222	5	23.8	66	22	ABB27016	Protein #9015 enco
223	5	23.8	66	22	AAM65157	Human brain expres
224	5	23.8	66	22	AAM77862	Human bone marrow
225	5	23.8	66	22	AAM21762	Peptide #8196 enco
226	5	23.8	66	22	AAM38084	Human brain expres
227	5	23.8	67	22	AAM94910	Peptide #12121 enc
228	5	23.8	67	22	AAO11655	Human reproductive
229	5	23.8	68	20	AAY07905	Human polypeptide
230	5	23.8	68	22	ABG00305	Novel human diagno

Human secreted pro

Novel human diagno

No.	Score		Pos	Len	Accession	Description
231	23.8	5	68	22	AAO13010	Human polypeptide
232	23.8	5	69	22	ABB30649	Peptide #3300 enco
233	23.8	5	69	22	ABB35820	Peptide #3326 enco
234	23.8	5	69	22	ABB21236	Protein #3235 enco
235	23.8	5	69	22	AAM69004	Human bone marrow
236	23.8	5	69	22	AAM90904	Human immune/haema
237	23.8	5	69	17	AAM29321	Peptide #3358 enco
238	23.8	5	70	22	AAR97850	Rat brain calcium
239	23.8	5	70	22	AAU50030	Propionibacterium
240	23.8	5	70	22	AAU59695	Propionibacterium
241	23.8	5	71	21	AAG18763	Zea mays protein f
242	23.8	5	71	21	AAG32668	Zea mays protein f
243	23.8	5	71	21	AAG37983	Arabidopsis thalia
244	23.8	5	71	21	AAG48400	Arabidopsis thalia
245	23.8	5	73	20	AAY73882	Human prostate tum
246	23.8	5	73	22	ABC27287	Novel human diagno
247	23.8	5	74	22	AAU56270	Propionibacterium
248	23.8	5	74	21	AAM39326	Human polypeptide
249	23.8	5	75	22	AAB08492	Amino acid sequenc
250	23.8	5	75	22	AAG67855	Murine leukemia vi
251	23.8	5	76	21	AAM84253	Human immune/haema
252	23.8	5	77	22	AAB34363	Human secreted pro
253	23.8	5	77	22	AAB28103	Human peptide #754
254	23.8	5	77	22	ABB33278	Peptide #784 encod
255	23.8	5	77	22	ABB18740	Protein #739 encod
256	23.8	5	77	22	AAM54070	Human brain expres
257	23.8	5	78	22	AAM66458	Human bone marrow
258	23.8	5	78	22	AAM14330	Peptide #764 encod
259	23.8	5	77	22	AAM26740	Peptide #777 encod
260	23.8	5	77	22	AAM02059	Peptide #741 encod
261	23.8	5	78	18	AAW23001	Canine herpesvirus
262	23.8	5	78	19	AAW72661	Canine herpes viru
263	23.8	5	78	22	AAY08239	Human cadherin-3 p
264	23.8	5	78	22	AAM82026	Human haematologic
265	23.8	5	78	22	AAM64235	Human brain expres
266	23.8	5	78	22	AAB51318	Canine herpes viru
267	23.8	5	79	21	AAB43069	Human ORFX ORF2833
268	23.8	5	79	22	AAG04436	Arabidopsis thalia
269	23.8	5	80	22	AAU23267	Novel human enzyme
270	23.8	5	81	22	AAU42097	Propionibacterium
271	23.8	5	81	22	AAU67375	Propionibacterium
272	23.8	5	82	13	AAR27371	TRFP Chain #2 C2ST
273	23.8	5	82	22	AAM23486	Human EST encoded
274	23.8	5	84	22	AAG02176	Human secreted pro
275	23.8	5	84	22	AAM89076	Human immune/haema
276	23.8	5	84	22	AAM06597	Human foetal prote
277	23.8	5	84	22	AAM06793	Human foetal prote
278	23.8	5	85	21	AAG37314	Arabidopsis thalia
279	23.8	5	85	22	AAG48399	Arabidopsis thalia
280	23.8	5	85	22	AAO07257	Human polypeptide
281	23.8	5	86	22	AAB50420	Human uncoupling p
282	23.8	5	86	21	AAU42401	Propionibacterium
283	23.8	5	86	22	AAU62283	Propionibacterium
284	23.8	5	86	21	ABG08617	Novel human diagno
285	23.8	5	87	22	AAG56639	Arabidopsis thalia
286	23.8	5	88	20	AAY73898	Human prostate tum
287	23.8	5	88	21	AAG37313	Arabidopsis thalia
288	23.8	5	88	22	AAU21374	Human novel foetal
289	23.8	5	88	22	AAM06338	Human foetal prote
290	23.8	5	88	22	AAB61901	Mouse apoptosis me
291	23.8	5	89	20	AAY29338	Human secreted pro
292	23.8	5	89	22	AAG36152	Arabidopsis thalia
293	23.8	5	89	22	AAU39089	Human secreted pro
294	23.8	5	89	21	AAO11912	Human polypeptide
295	23.8	5	90	22	AAG02990	Human secreted pro
296	23.8	5	90	22	AAU14264	Human novel protei
297	23.8	5	90	22	AAU14500	Human novel protei
298	23.8	5	91	21	AAG18762	Zea mays protein f
299	23.8	5	91	21	AAG37981	Arabidopsis thalia
300	23.8	5	91	21	ABG21730	Novel human diagno
301	23.8	5	91	22	AAM85962	Human immune/haema
302	23.8	5	92	15	AAR65454	T-cell receptor V-
303	23.8	5	92	20	AAY37880	Chlamydia trachoma
304	23.8	5	92	22	ABG08059	Novel human diagno
305	23.8	5	93	22	AAU56795	Propionibacterium
306	23.8	5	93	22	ABG19225	Novel human diagno
307	23.8	5	94	19	AAW79096	Human secreted pro
308	23.8	5	94	21	AAB57028	Human prostate can
309	23.8	5	94	22	AAG55575	Arabidopsis thalia
310	23.8	5	94	21	ABG28213	Novel human diagno
311	23.8	5	95	18	AAW28022	Amino acid sequenc
312	23.8	5	95	20	AAY59742	Human normal ovari
313	23.8	5	95	21	AAG60424	Arabidopsis thalia
314	23.8	5	96	22	AAU56723	Propionibacterium
315	23.8	5	97	12	AAR12123	TRFP chain 2 - tru
316	23.8	5	97	18	AAY11247	Streptococcus pneu
317	23.8	5	97	19	AAW38634	Streptococcus pneu
318	23.8	5	97	21	AAB28935	T cell reactive fe
319	23.8	5	97	21	AAY87675	Feline human TRFP
320	23.8	5	97	21	AAY90105	Cat TRFP chain 2 t
321	23.8	5	97	21	AAY51472	Human TRFP chain 2
322	23.8	5	97	22	AAU45726	Propionibacterium
323	23.8	5	97	22	AAU46334	Propionibacterium
324	23.8	5	97	22	AAM01050	CFE 53 protein seq
325	23.8	5	98	22	AAU67006	Propionibacterium
326	23.8	5	98	22	AAU42269	Propionibacterium
327	23.8	5	98	22	AAU55100	Arabidopsis thalia
328	23.8	5	100	21	AAG16261	Arabidopsis thalia
329	23.8	5	100	22	AAM81404	Human haematologic
330	23.8	5	100	22	AAM81907	Human haematologic
331	23.8	5	101	13	AAR27372	TRFP Chain #2 C2ST
332	23.8	5	101	21	AAB57154	Human prostate can
333	23.8	5	101	22	AAB42276	Human ORFX ORF2040
334	23.8	5	101	22	AAU40192	Propionibacterium
335	23.8	5	101	22	AAU06242	Propionibacterium
336	23.8	5	102	22	AAM86328	Human immune/haema
337	23.8	5	102	22	AAG74435	Human colon cancer
338	23.8	5	102	20	AAY12429	Human 5' EST secre
339	23.8	5	103	21	AAB40704	Human ORFX ORF468
340	23.8	5	103	22	AAM81878	Human haematologic
341	23.8	5	103	22	AAM81942	Human haematologic
342	23.8	5	104	15	AAR65453	T-cell receptor V-
343	23.8	5	104	21	AAG24831	Arabidopsis thalia
344	23.8	5	105	20	AAY37921	Chlamydia trachoma
345	23.8	5	105	22	AAM63587	Human brain expres
346	23.8	5	105	22	AAM76402	Human bone marrow
347	23.8	5	105	22	AAM36511	Peptide #10548 enc
348	23.8	5	105	22	AAG62631	C sorokiniana omeg
349	23.8	5	106	12	AAR15548	Asn 9 pGRF precurs
350	23.8	5	106	12	AAR15449	Ser 9 pGRF precurs
351	23.8	5	106	20	AAY23520	Amino acid sequenc
352	23.8	5	106	22	AAU50536	Propionibacterium
353	23.8	5	106	22	AAB47452	Human polypeptide
354	23.8	5	106	22	AAB47453	Fab' CH1. Synthet
355	23.8	5	106	22	AAG63351	Fab', H-N162D. Syn
356	23.8	5	106	22	AAG63352	Amino acid sequenc
357	23.8	5	107	21	AAB41840	Human ORFX ORF1604
358	23.8	5	107	21	AAG26469	Arabidopsis thalia
359	23.8	5	107	21	AAG47776	Arabidopsis thalia
360	23.8	5	107	21	AAG48382	Arabidopsis thalia
361	23.8	5	108	21	AAG44113	Arabidopsis thalia
362	23.8	5	109	22	AAR12122	TRFP I chain 2 - s
363	23.8	5	109	12	AAR36541	TRFP chain 2 (with
364	23.8	5	109	14	AAR41985	Human T cell react
365	23.8	5	109	14	AAY25593	Felis sp. allergen
366	23.8	5	109	20	AAB28934	T cell reactive fe
367	23.8	5	109	21	AAY87674	Feline human TRFP
368	23.8	5	109	21	AAY90104	Cat TRFP chain 2 s
369	23.8	5	109	21	AAY51471	Human TRFP chain 2
370	23.8	5	109	21	AAU02024	B.thuringiensis t
371	23.8	5	110	22	AAR27370	TRFP Chain #2 with
372	23.8	5	110	13	AAY37848	Chlamydia trachoma
373	23.8	5	110	20	AAG47775	Arabidopsis thalia
374	23.8	5	111	21	AAR12121	TRFP chain 2 - lon
375	23.8	5	111	12	AAR22369	TRFP Chain #2 with
376	23.8	5	111	13	AAB24483	Human ORFX ORF2247

No.			Len	Accession	ID	Description
377	5	23.8	111	AAB28933	21	T cell reactive fe
378	5	23.8	111	AAY87673	21	Feline human TRFP
379	5	23.8	111	AAY90103	21	Cat TRFP chain 2 1
380	5	23.8	111	AAY51470	21	Human TRFP chain 2
381	5	23.8	111	AAO06213	22	Human polypeptide
382	5	23.8	112	AAU43924	22	Propionibacterium
383	5	23.8	112	AAU45902	22	Propionibacterium
384	5	23.8	113	AAU10210	22	Human polypeptide
385	5	23.8	113	AAU02025	22	B. thuringiensis t
386	5	23.8	114	AAU61082	22	Propionibacterium
387	5	23.8	116	AAP70305	8	Sequence of polype
388	5	23.8	116	AAR90241	16	Interleukin-1beta
389	5	23.8	116	AAU49642	22	Propionibacterium
390	5	23.8	117	AAB60517	22	Human des-Gln14-gh
391	5	23.8	117	AAW87991	20	Protein designated
392	5	23.8	117	AAG18008	21	Arabidopsis thalia
393	5	23.8	117	AAG24441	22	Arabidopsis thalia
394	5	23.8	117	AAG47756	21	Arabidopsis thalia
395	5	23.8	117	AAY87236	21	Human signal pepti
396	5	23.8	117	AAO10712	22	Human polypeptide
397	5	23.8	117	AAM38890	22	Human polypeptide
398	5	23.8	117	AAB62649	22	Human zsig33 polyp
399	5	23.8	117	AAB20101	22	Zsig33 protein. H
400	5	23.8	117	AAB60511	22	Human ghrelin prep
401	5	23.8	118	AAY66708	21	Membrane-bound pro
402	5	23.8	118	AAM51702	22	Human tubulin fold
403	5	23.8	118	AAU30705	22	Novel human secret
404	5	23.8	118	AAO01659	22	Human polypeptide
405	5	23.8	118	AAU16396	22	Human novel secret
406	5	23.8	118	AAU12392	22	Human PRO1066 poly
407	5	23.8	118	AAG77380	22	Human colon cancer
408	5	23.8	118	AAB65231	22	Human PRO1066 (UNQ
409	5	23.8	119	AAG18007	21	Arabidopsis thalia
410	5	23.8	119	AAM81204	22	Human haematologic
411	5	23.8	120	AAR62045	15	T. ovis vaccine ca
412	5	23.8	120	AAR92832	17	Taenia ovis antige
413	5	23.8	120	ABG27167	22	Novel human diagno
414	5	23.8	120	AAU49599	22	Novel human diagno
415	5	23.8	120	AAO01129	22	Human reproductive
416	5	23.8	120	AAO01846	22	Human polypeptide
417	5	23.8	120	AAO10647	22	Human polypeptide
418	5	23.8	121	AAU50432	22	Propionibacterium
419	5	23.8	121	AAM23530	22	Human EST encoded
420	5	23.8	122	AAR54101	15	Humanised MAb H-ch
421	5	23.8	123	AAR56651	15	Anti-CMV monoclona
422	5	23.8	123	ABB70162	22	Drosophila melanog
423	5	23.8	123	ABG12707	22	Novel human diagno
424	5	23.8	123	ABB27562	22	Human peptide #213
425	5	23.8	123	ABB32721	22	Peptide #227 encod
426	5	23.8	123	ABB18207	22	Protein #206 encod
427	5	23.8	123	ABB12232	22	Human secreted pro
428	5	23.8	123	AAM53539	22	Human reproductive
429	5	23.8	123	AAM65920	22	Human brain expres
430	5	23.8	123	AAM13787	22	Human bone marrow
431	5	23.8	123	AAM26185	22	Peptide #221 encod
432	5	23.8	123	AAM01532	22	Peptide #222 encod
433	5	23.8	123	AAW60168	22	Peptide #214 encod
434	5	23.8	125	AAU31890	19	Light chain variab
435	5	23.8	125	AAW60167	22	Novel human secret
436	5	23.8	126	AAB41342	19	Light chain variab
437	5	23.8	126	AAB25791	21	Human ORFX ORF1106
438	5	23.8	126	AAY65398	21	Human secreted pro
439	5	23.8	126	ABG25140	22	Human 5' EST relat
440	5	23.8	126	AAM82236	22	Novel human diagno
441	5	23.8	126	AAO00851	22	Human immune/haema
442	5	23.8	126	AAM40676	22	Human polypeptide
443	5	23.8	126	AAB75377	22	Human polypeptide
444	5	23.8	127	AAU60961	21	Human secreted pro
445	5	23.8	128	AAM50194	22	Propionibacterium
446	5	23.8	128	AAU45066	22	Propionibacterium
447	5	23.8	129	AAG41531	21	Arabidopsis thalia
448	5	23.8	129	AAG46112	21	Arabidopsis thalia
449	5	23.8	129	AAU57573	22	Propionibacterium
450	5	23.8	129	AAM81819	22	Human haematologic
451	5	23.8	130	AAW61553	19	Hyalophora moth gl
452	5	23.8	130	AAB43419	21	Human cancer assoc
453	5	23.8	130	AAY94298	21	Giant silk moth an
454	5	23.8	131	ABG29896	22	Novel human diagno
455	5	23.8	131	AAG91450	22	C glutamicum prote
456	5	23.8	133	ABB12201	22	Human IGERB homolo
457	5	23.8	133	AAB80035	22	Corynebacterium gl
458	5	23.8	134	ABB61057	22	Drosophila melanog
459	5	23.8	135	AAU32061	22	Novel human secret
460	5	23.8	136	AAE05855	22	Pseudomonas stutze
461	5	23.8	136	AAB45111	21	Human secreted pro
462	5	23.8	138	ABG20761	22	Novel human diagno
463	5	23.8	141	AAY30917	20	H. tuberosus lecti
464	5	23.8	143	AAG23325	21	Arabidopsis thalia
465	5	23.8	145	AAG56338	21	Arabidopsis thalia
466	5	23.8	145	AAO07941	22	Human polypeptide
467	5	23.8	145	AAG75023	22	Human colon cancer
468	5	23.8	145	AAB80036	22	Corynebacterium gl
469	5	23.8	146	AAG36291	21	Arabidopsis thalia
470	5	23.8	146	AAG89776	22	C glutamicum prote
471	5	23.8	147	AAM80075	22	Human haematologic
472	5	23.8	147	AAM81332	22	Human haematologic
473	5	23.8	147	AAM81886	22	Human haematologic
474	5	23.8	147	AAU20581	22	Human secreted pro
475	5	23.8	149	ABG17274	22	Novel human diagno
476	5	23.8	149	ABG17960	22	Novel human diagno
477	5	23.8	149	AAO04215	22	Human polypeptide
478	5	23.8	150	AAG01651	21	Human secreted pro
479	5	23.8	150	AAM94552	22	Human reproductive
480	5	23.8	151	AAW54438	22	Mouse novel secret
481	5	23.8	151	AAB10245	21	Murine adult splee
482	5	23.8	151	AAG18840	22	Arabidopsis thalia
483	5	23.8	151	AAG21713	21	Arabidopsis thalia
484	5	23.8	151	AAG51027	21	Arabidopsis thalia
485	5	23.8	151	AAM25916	22	Human protein sequ
486	5	23.8	151	AAM42142	22	Human polypeptide
487	5	23.8	152	ABG26000	22	Novel human diagno
488	5	23.8	152	AAG90353	22	C glutamicum prote
489	5	23.8	154	AAU66436	22	Novel human diagno
490	5	23.8	154	ABG03121	22	Propionibacterium
491	5	23.8	155	AAB40744	21	Human ORFX ORF508
492	5	23.8	156	AAY97382	21	Spinach mature lum
493	5	23.8	156	AAU45177	22	Propionibacterium
494	5	23.8	156	AAU02202	22	Retinol-binding pr
495	5	23.8	156	AAB04149	22	Spinach lumazine s
496	5	23.8	157	AAU29455	22	Human G protein-co
497	5	23.8	158	ABG27299	22	Novel human diagno
498	5	23.8	158	AAG75714	22	Human colon cancer
499	5	23.8	159	AAU57541	22	Propionibacterium
500	5	23.8	160	AAY70542	21	Protein encoded by
501	5	23.8	160	ABB67163	22	Drosophila melanog
502	5	23.8	160	AAM94960	22	Human reproductive
503	5	23.8	161	AAE00216	22	Murine N-terminal
504	5	23.8	162	AAB38478	21	Fragment of human
505	5	23.8	162	AAY70540	22	Protein encoded by
506	5	23.8	162	ABG15181	22	Novel human diagno
507	5	23.8	163	AAU21111	22	Novel human diagno
508	5	23.8	164	AAU60868	21	Propionibacterium
509	5	23.8	164	AAG37312	22	Arabidopsis thalia
510	5	23.8	165	AAG81348	22	Human AFP protein
511	5	23.8	165	AAU57040	22	Propionibacterium
512	5	23.8	166	AAG75335	22	Human colon cancer
513	5	23.8	166	AAG09254	22	Arabidopsis thalia
514	5	23.8	166	AAG18339	21	Arabidopsis thalia
515	5	23.8	166	ABB67726	22	Drosophila melanog
516	5	23.8	166	AAE09787	22	Paenibacillus pabu
517	5	23.8	167	AAY53632	21	A bone marrow secr
518	5	23.8	167	ABG17123	22	Novel human diagno
519	5	23.8	167	AAU01006	22	Tomato phosphoenol
520	5	23.8	168	AAU01007	22	Cotton phosphoenol
521	5	23.8	168	AAG09253	21	Arabidopsis thalia
522	5	23.8	168	AAG09997	21	Arabidopsis thalia

No.	Score		Pos	Len	Accession	Description
523	23.8	5	168	21	AAG36151	Arabidopsis thalia
524	23.8	5	168	21	AAG37199	Arabidopsis thalia
525	23.8	5	168	21	AAG38540	Arabidopsis thalia
526	23.8	5	170	19	AAW64200	Human calcineurin.
527	23.8	5	170	21	AAB09977	Human CNBII protei
528	23.8	5	170	21	AAB09978	Human HCNB protein
529	23.8	5	170	21	AAG09252	Arabidopsis thalia
530	23.8	5	170	21	AAG09996	Arabidopsis thalia
531	23.8	5	170	21	AAG37198	Arabidopsis thalia
532	23.8	5	170	21	AAG38539	Arabidopsis thalia
533	23.8	5	172	18	AAW26532	Trypanosoma cruzi
534	23.8	5	172	20	AAY23294	Trpanosoma cruzi a
535	23.8	5	173	22	AAU23564	Novel human enzyme
536	23.8	5	173	22	AAB64410	Amino acid sequenc
537	23.8	5	174	22	AAE09790	Paenibacillus pabu
538	23.8	5	176	22	AAG36150	Arabidopsis thalia
539	23.8	5	176	22	AAB50780	Human secreted pro
540	23.8	5	177	21	AAG48381	Arabidopsis thalia
541	23.8	5	178	22	AAE12073	Dendritic cell (DC
542	23.8	5	179	19	AAW40068	Human monoclonal a
543	23.8	5	179	21	AAG38559	Arabidopsis thalia
544	23.8	5	179	23	AAE00045	Human monoclonal a
545	23.8	5	181	21	AAG11003	Arabidopsis thalia
546	23.8	5	181	21	AAG47970	Arabidopsis thalia
547	23.8	5	181	21	AAG47972	Arabidopsis thalia
548	23.8	5	181	21	ABG11441	Novel human diagno
549	23.8	5	181	22	AAG64378	Human PDZ protein
550	23.8	5	182	22	ABG19219	Novel human diagno
551	23.8	5	183	22	ABB27546	Human peptide #197
552	23.8	5	183	22	ABB32705	Peptide #211 encod
553	23.8	5	183	22	ABB18194	Protein #193 encod
554	23.8	5	183	22	AAM53526	Human brain expres
555	23.8	5	183	22	AAM65905	Human bone marrow
556	23.8	5	183	22	AAM83966	Human immune/haema
557	23.8	5	183	22	AAM13772	Peptide #206 encod
558	23.8	5	183	22	AAM26170	Peptide #207 encod
559	23.8	5	183	22	AAM01517	Peptide #199 encod
560	23.8	5	184	20	AAY37088	Amino acid sequenc
561	23.8	5	185	22	ABB12284	Human secreted pro
562	23.8	5	185	22	AAM80069	Human protein SEQ
563	23.8	5	186	22	AAG38538	Arabidopsis thalia
564	23.8	5	186	21	AAG38701	Arabidopsis thalia
565	23.8	5	186	22	AAG47774	Arabidopsis thalia
566	23.8	5	186	21	AAG48380	Arabidopsis thalia
567	23.8	5	186	22	ABB66486	Drosophila melanog
568	23.8	5	187	18	AAY11193	S. pneumoniae phos
569	23.8	5	187	20	AAY59947	Human endometrium
570	23.8	5	187	22	AAY72771	Saccharomyces cere
571	23.8	5	188	21	AAY95043	Candida albicans p
572	23.8	5	189	22	AAM95239	Human reproductive
573	23.8	5	189	22	AAM43564	Human polypeptide
574	23.8	5	189	22	AAM43639	Human polypeptide
575	23.8	5	189	22	AAU19951	Novel human calciu
576	23.8	5	192	22	AAW74885	Human secreted pro
577	23.8	5	192	21	AAY91681	Human secreted pro
578	23.8	5	193	7	AAP61520	Sequence of new an
579	23.8	5	193	22	ABG08106	Novel human diagno
580	23.8	5	193	22	ABG20816	Novel human diagno
581	23.8	5	193	22	ABB53068	Escherichia coli p
582	23.8	5	194	21	AAG09995	Arabidopsis thalia
583	23.8	5	194	22	AAG37197	Arabidopsis thalia
584	23.8	5	195	20	AAY26185	Corn 2-isopropylma
585	23.8	5	196	15	AAR43986	pIse Protein. Hum
586	23.8	5	196	15	AAR55343	Sequence of envelo
587	23.8	5	199	7	AAP61521	Sequence of new an
588	23.8	5	199	22	ABB28364	Peptide #1015 enco
589	23.8	5	199	22	ABB33543	Peptide #1049 enco
590	23.8	5	199	22	ABB18999	Protein #998 encod
591	23.8	5	199	22	AAM54319	Human brain expres
592	23.8	5	199	22	AAM66717	Human bone marrow
593	23.8	5	199	22	AAM14582	Peptide #1016 enco
594	23.8	5	199	22	AAM27003	Peptide #1040 enco
595	23.8	5	199	22	AAM02308	Peptide #990 encod
596	23.8	5	200	18	AAW55524	H. pylori ORF 29ep
597	23.8	5	201	21	AAG47971	Arabidopsis thalia
598	23.8	5	201	21	AAG11002	Arabidopsis thalia
599	23.8	5	201	22	AAU19291	Human G protein-co
600	23.8	5	201	14	AAR52020	Truncated human In
601	23.8	5	202	21	AAG05783	Arabidopsis thalia
602	23.8	5	203	21	AAG49184	Arabidopsis thalia
603	23.8	5	203	16	AAR59481	Prolactin peptide
604	23.8	5	204	21	AAG22337	Arabidopsis thalia
604	23.8	5	204	21	AAY91396	Human secreted pro
606	23.8	5	205	15	AAR52660	Porphyrin antibody
607	23.8	5	205	21	AAB25288	Eucalyptus grandis
608	23.8	5	205	22	AAB61900	Mouse apoptosis me
609	23.8	5	208	21	AAG37048	Arabidopsis thalia
610	23.8	5	209	21	AAG47969	Arabidopsis thalia
611	23.8	5	210	22	ABB66738	Drosophila melanog
612	23.8	5	210	22	ABG06390	Novel human diagno
613	23.8	5	210	22	AAU25652	G protein-coupled
614	23.8	5	211	22	ABB61507	Drosophila melanog
615	23.8	5	212	17	AAW15932	Antibody 7G12 heav
616	23.8	5	213	16	AAR64201	Monoclonal antibod
617	23.8	5	213	21	AAG14667	Arabidopsis thalia
618	23.8	5	214	20	AAY44176	MAb Fab13B5 heavy
619	23.8	5	215	16	AAR64203	Monoclonal antibod
620	23.8	5	215	22	AAM41747	Human polypeptide
621	23.8	5	216	22	AAU03676	Group B Streptococ
622	23.8	5	217	15	AAR53805	FAB heavy chain fo
623	23.8	5	217	16	AAR75460	Mouse antibody H3-
624	23.8	5	217	21	AAG18006	Arabidopsis thalia
625	23.8	5	219	19	AAW53597	Amino acid sequenc
626	23.8	5	220	15	AAR74782	Heavy chain of 59.
627	23.8	5	221	21	AAB63215	Gene 39 human secr
628	23.8	5	221	15	AAG01424	Human secreted pro
629	23.8	5	222	15	AAR59415	6D9 antibody heavy
630	23.8	5	222	17	AAW15934	Antibody 3G2 heavy
631	23.8	5	222	22	AAY97388	Spinach precursor
632	23.8	5	222	22	AAG92413	C glutanicum prote
633	23.8	5	222	22	AAB04155	Spinach lumazine s
634	23.8	5	223	16	AAR75456	Mouse antibody FB3
635	23.8	5	223	16	AAR75458	Mouse antibody F4-
636	23.8	5	225	21	AAG22336	Arabidopsis thalia
637	23.8	5	225	22	AAY94449	Human inflammation
638	23.8	5	226	22	AAG65272	Human cell surface
639	23.8	5	226	22	ABB67771	Drosophila melanog
640	23.8	5	226	22	AAB98706	Chandra, a helper
641	23.8	5	227	18	AAW55386	H. pylori ORF 1lge
642	23.8	5	227	22	ABG19236	Novel human diagno
643	23.8	5	228	21	AAB34732	Human secreted pro
644	23.8	5	228	22	AAU18542	Human cytoskeletal
645	23.8	5	229	21	AAG38700	Arabidopsis thalia
646	23.8	5	230	18	AAM27088	Mouse monoclonal a
647	23.8	5	230	21	AAB32942	Pinus radiata tran
648	23.8	5	230	22	ABG11220	Novel human diagno
649	23.8	5	231	18	AAW27090	Mouse monoclonal a
650	23.8	5	231	21	AAG22134	Arabidopsis thalia
651	23.8	5	232	22	AAU48673	Propionibacterium
652	23.8	5	233	22	AAY69007	Heavy chain of an
653	23.8	5	236	15	AAR58705	Apo-B RNA editing
654	23.8	5	236	21	AAG22335	Arabidopsis thalia
655	23.8	5	236	22	AAE04574	Human G-protein co
656	23.8	5	238	18	AAP60307	Sequence of new an
657	23.8	5	238	8	AAP70306	Sequence of interl
658	23.8	5	238	14	AAR42447	Human interleukin-
658	23.8	5	238	19	AAW23831	S. glaucescens acb
659	23.8	5	238	22	ABG15868	Novel human diagno
661	23.8	5	238	22	AAM25683	Human protein sequ
662	23.8	5	239	22	AAY43287	HFSH beta subunit-
663	23.8	5	239	20	AAY43294	HFSH beta subunit-
664	23.8	5	239	20	AAW96745	High affinity immu
665	23.8	5	239	21	AAY50174	Human high affinit
666	23.8	5	239	22	ABG15553	Novel human diagno
667	23.8	5	239	22	AAB74584	Human HAIERbs isof
668	23.8	5	240	20	AAY43288	HTSH beta subunit-

Num	Score		Pos	Len	Accession	Description
669	23.8	5	240	20	AAV43295	HTSH beta subunit-
670	23.8	5	240	21	AAG54898	Arabidopsis thalia
671	23.8	5	240	21	AAY69004	Light chain of an
672	23.8	5	241	21	AAG32098	Arabidopsis thalia
673	23.8	5	242	20	AAY43286	HLH beta subunit-J
674	23.8	5	242	20	AAY43293	HLH beta subunit-J
675	23.8	5	242	21	AAG10359	Arabidopsis thalia
676	23.8	5	242	21	AAG20310	Arabidopsis thalia
677	23.8	5	242	21	AAG29611	Arabidopsis thalia
678	23.8	5	242	21	AAG47742	Arabidopsis thalia
679	23.8	5	242	21	AAG47751	Arabidopsis thalia
680	23.8	5	242	21	AAG49913	Arabidopsis thalia
681	23.8	5	243	22	AAY73495	Human secreted pro
682	23.8	5	245	21	ABG01294	Arabidopsis thalia
683	23.8	5	245	21	AAG10123	Arabidopsis thalia
684	23.8	5	245	22	AAY94973	Human secreted pro
685	23.8	5	245	22	ABG14073	Novel human diagno
686	23.8	5	246	21	AAG38699	Arabidopsis thalia
687	23.8	5	246	21	AAG49984	Arabidopsis thalia
688	23.8	5	247	16	AAW11917	Murine MAb SK48-E2
689	23.8	5	247	21	AAG10122	Arabidopsis thalia
690	23.8	5	247	21	AAG29610	Arabidopsis thalia
691	23.8	5	247	21	AAG47741	Arabidopsis thalia
692	23.8	5	247	22	ABG20214	Novel human diagno
693	23.8	5	248	18	AAW20921	H. pylori derived
694	23.8	5	248	20	AAY15225	Human receptor pro
695	23.8	5	248	20	AAY36046	Extended human sec
696	23.8	5	248	21	AAY91531	Human secreted pro
697	23.8	5	248	22	AAE12072	Dendritic cell (DC
698	23.8	5	248	22	AAB70489	Human hHAIERbs-iso
699	23.8	5	249	21	AAB74583	Human HAIERbs isof
700	23.8	5	250	17	AAW06177	Murine A5B57 Fd fr
701	23.8	5	250	19	AAW61619	Clone HTPEF86 of T
702	23.8	5	250	20	AAY48505	Human breast tumou
703	23.8	5	250	21	AAB58800	Breast and ovarian
704	23.8	5	250	21	AAB58800	Breast and ovarian
705	23.8	5	250	21	AAY70432	Human cell surface
706	23.8	5	250	22	ABB71004	Drosophila melanog
707	23.8	5	250	22	ABG08618	Novel human diagno
708	23.8	5	251	22	AAU43858	Propionibacterium
709	23.8	5	252	22	AAB92903	Human protein sequ
710	23.8	5	252	22	AAB93630	Human protein sequ
711	23.8	5	253	10	AAP94686	Amino acid sequenc
712	23.8	5	254	20	AAY48226	Human prostate can
713	23.8	5	254	20	AAY30117	Murine anti-botuli
714	23.8	5	254	20	AAY30119	Murine anti-botuli
715	23.8	5	254	20	AAY30121	Murine anti-botuli
716	23.8	5	254	20	AAY30123	Murine anti-botuli
717	23.8	5	254	22	AAG78262	Human PTP1B. Homo
718	23.8	5	254	22	AAB47091	A-chain isoform fo
719	23.8	5	254	21	AAB59364	Human protein tyro
720	23.8	5	256	22	ABB69509	Drosophila melanog
721	23.8	5	257	21	AAB58419	Lung cancer associ
722	23.8	5	257	21	AAG20916	Arabidopsis thalia
723	23.8	5	257	21	AAG37992	Arabidopsis thalia
724	23.8	5	258	22	AAG47237	Arabidopsis thalia
725	23.8	5	258	22	ABB67761	Drosophila melanog
726	23.8	5	259	21	AAG47750	Arabidopsis thalia
727	23.8	5	259	21	AAG58988	Arabidopsis thalia
728	23.8	5	259	21	AAG59464	Arabidopsis thalia
729	23.8	5	260	21	AAG47236	Arabidopsis thalia
730	23.8	5	260	21	AAY95257	WOW-1 Fab heavy ch
731	23.8	5	262	22	AAB64618	Drosophila melanog
732	23.8	5	263	22	ABG07958	Novel human diagno
733	23.8	5	263	22	AAB24127	Novel human diagno
734	23.8	5	264	21	AAB32088	Human secreted pro
735	23.8	5	264	22	ABG18541	Novel human diagno
736	23.8	5	265	21	AAG10358	Arabidopsis thalia
737	23.8	5	265	22	ABG04256	Novel human diagno
738	23.8	5	265	22	ABG27172	Novel human diagno
739	23.8	5	266	22	ABB68856	Drosophila melanog
740	23.8	5	267	19	AAW61120	URA3 protein seque
741	23.8	5	267	20	AAY29831	Saccharomyces cere
742	23.8	5	267	20	AAY15224	Human receptor pro
743	23.8	5	267	20	AAY16789	S. cerevisiae URA3
744	23.8	5	267	21	AAY74716	Neisseria gonorrhe
745	23.8	5	267	21	AAY74719	Neisseria gonorrhe
746	23.8	5	267	22	AAG78230	Yeast URA3. Sacch
747	23.8	5	269	6	AAP50043	Sequence of interl
748	23.8	5	269	9	AAP81197	Interleukin-1 of h
749	23.8	5	269	12	AAR15747	Interleukin-1 beta
750	23.8	5	269	12	AAR15748	Interleukin-1 beta
751	23.8	5	269	14	AAR42213	Human Interleukin-
752	23.8	5	269	15	AAR46337	PB92 serine protea
753	23.8	5	269	15	AAR46338	Subtilisin 309 Ser
754	23.8	5	269	15	AAR46339	PB92 serine protea
755	23.8	5	269	15	AAR46340	Subtilisin 309 Ser
756	23.8	5	269	15	AAR46341	PB92 serine protea
757	23.8	5	269	15	AAR46342	Subtilisin 309 Ser
758	23.8	5	269	15	AAR46343	PB92 serine protea
759	23.8	5	269	15	AAR46344	Subtilisin 309 Ser
760	23.8	5	269	18	AAW31668	Precursor form of
761	23.8	5	269	19	AAW71382	Interleukin-1beta
762	23.8	5	269	20	AAY08322	Human IL-1beta pro
763	23.8	5	269	21	AAB37790	Human interleukin-
764	23.8	5	269	21	AAY96877	Human precursor in
765	23.8	5	269	21	AAY69531	Human interleukin
766	23.8	5	269	22	AAB97079	Human IL1B. Homo
767	23.8	5	269	22	AAB35251	Human pre-interleu
768	23.8	5	271	22	AAU25560	Human G Protein-Co
769	23.8	5	272	22	ABG16055	Novel human diagno
770	23.8	5	273	20	AAY43285	HCG beta subunit-J
771	23.8	5	273	20	AAY43289	HCG/hFSH beta subu
772	23.8	5	273	20	AAY43290	HCG/HTSH beta subu
773	23.8	5	273	20	AAY43292	HCG beta subunit-J
774	23.8	5	273	20	AAY43296	HCG/hFSH beta subu
775	23.8	5	273	20	AAY43297	HCG/HTSH beta subu
776	23.8	5	273	21	AAY91680	Human secreted pro
777	23.8	5	273	22	ABG17004	Novel human diagno
778	23.8	5	273	22	ABG25381	Novel human diagno
779	23.8	5	273	22	ABB71989	Human secreted pro
780	23.8	5	273	22	AAU32712	Novel human secret
781	23.8	5	273	22	AAM25809	Human protein sequ
782	23.8	5	273	22	AAG75619	Human colon cancer
783	23.8	5	274	22	ABG09578	Novel human diagno
784	23.8	5	275	22	ABG25380	Novel human diagno
785	23.8	5	276	21	AAG58018	Arabidopsis thalia
786	23.8	5	276	22	ABB69618	Drosophila melanog
787	23.8	5	276	22	AAU55024	Propionibacterium
788	23.8	5	276	22	ABG15182	Novel human diagno
789	23.8	5	276	22	ABG15285	Novel human diagno
790	23.8	5	276	22	AAB71708	Novel human diagno
791	23.8	5	278	22	ABG15466	Novel human diagno
792	23.8	5	279	22	ABG11440	Novel human diagno
793	23.8	5	279	22	AAU01011	Human secreted pro
794	23.8	5	280	21	AAB24485	Human ORFX ORF2249
795	23.8	5	282	16	AAR71460	Human surfactant p
796	23.8	5	282	22	ABG02219	Novel human diagno
797	23.8	5	283	22	ABB61572	Drosophila melanog
798	23.8	5	283	22	AAU38430	Salmonella typhi c
799	23.8	5	283	22	AAB99521	Saccharomyces cere
800	23.8	5	284	22	ABB59014	Drosophila melanog
801	23.8	5	286	22	ABB71833	Drosophila melanog
802	23.8	5	287	17	AAW06967	Cyclopropane mycol
803	23.8	5	287	19	AAW40071	Human H11-scFv con
804	23.8	5	287	22	ABG19235	Novel human diagno
805	23.8	5	287	22	AAE00948	Human monoclonal a
806	23.8	5	289	22	AAG58017	Arabidopsis thalia
807	23.8	5	289	22	ABG23002	Novel human diagno
808	23.8	5	290	22	AAO05926	Human polypeptide
809	23.8	5	291	22	ABB69304	Drosophila melanog
810	23.8	5	291	22	ABB10444	Human cDNA SEQ ID
811	23.8	5	291	22	AAU35800	Novel human enzyme
812	23.8	5	292	21	AAG10573	Arabidopsis thalia
813	23.8	5	292	21	AAG32097	Arabidopsis thalia
814	23.8	5	292	21	AAG46692	Arabidopsis thalia

#					Accession	Description
815	5	23.8	293	21	AAG47235	Arabidopsis thalia
816	5	23.8	293	21	AAG58016	Arabidopsis thalia
817	5	23.8	294	21	AAG22915	Arabidopsis thalia
818	5	23.8	294	21	AAG37991	Arabidopsis thalia
819	5	23.8	295	7	AAP60326	Interleukin-1 gene
820	5	23.8	295	7	AAP60680	Prepro- and mature
821	5	23.8	295	12	AAR14855	Mature Interleukin
822	5	23.8	295	21	AAG20914	Arabidopsis thalia
823	5	23.8	295	21	AAG37990	Arabidopsis thalia
824	5	23.8	295	22	ABB52467	Escherichia coli p
825	5	23.8	296	22	AAU37340	Staphyloccocus aur
826	5	23.8	296	22	AAM25235	Human protein sequ
827	5	23.8	296	22	AAG82937	S. epidermidis ope
828	5	23.8	297	17	AAW00283	Mutant farnesyldip
829	5	23.8	297	17	AAW00286	Native farnesyldip
830	5	23.8	298	21	AAG84912	Shrimp white spot
831	5	23.8	299	21	AAG05420	Arabidopsis thalia
832	5	23.8	299	21	AAG47773	Arabidopsis thalia
833	5	23.8	299	21	AAY91352	Human secreted pro
834	5	23.8	300	20	AAY41759	Human PRO1014 prot
835	5	23.8	300	21	AAB44315	Human PRO1014 (UNQ
836	5	23.8	300	21	AAU22928	Novel human enzyme
837	5	23.8	300	22	AAU29089	Human PRO polypept
838	5	23.8	300	22	AAB19930	Human oxidoreducta
839	5	23.8	301	21	AAB18895	A maize chitinase
840	5	23.8	301	21	AAG14666	Arabidopsis thalia
841	5	23.8	302	21	AAB54312	Human pancreatic c
842	5	23.8	302	21	AAG49983	Arabidopsis thalia
843	5	23.8	302	22	AAU16378	Human novel secret
844	5	23.8	302	22	AAG90217	C glutamicum prote
845	5	23.8	303	18	AAW24590	H. pylori derived
846	5	23.8	303	18	AAW20132	H. Pylori derived
847	5	23.8	303	21	AAG10121	Arabidopsis thalia
848	5	23.8	303	22	AAG90813	C glutamicum prote
849	5	23.8	303	22	AAB87805	Mouse T2R11 amino
850	5	23.8	304	19	AAW40070	Human H11-scFv con
851	5	23.8	304	22	AAU07621	Mouse 2P channel p
852	5	23.8	304	22	AAE00947	Human monoclonal a
853	5	23.8	305	22	AAU37698	Streptococcus pneu
854	5	23.8	305	22	AAU37960	Streptococcus pneu
855	5	23.8	307	21	AAY91421	Human secreted pro
856	5	23.8	307	22	AAU33659	Pseudomonas aerugi
857	5	23.8	308	21	AAG37210	Arabidopsis thalia
858	5	23.8	309	21	AAG05419	Arabidopsis thalia
859	5	23.8	309	21	AAG47772	Arabidopsis thalia
860	5	23.8	309	22	AAU38118	Salmonella typhi C
861	5	23.8	310	22	ABG05624	Novel human diagno
862	5	23.8	311	20	AAW90248	Human FAST-1 prote
863	5	23.8	311	21	AAG37209	Arabidopsis thalia
864	5	23.8	311	22	ABB71370	Drosophila melanog
865	5	23.8	313	21	AAB14138	Bordetella pertuss
866	5	23.8	313	22	AAY91459	Human secreted pro
867	5	23.8	313	22	AAY71941	Human olfactory re
868	5	23.8	314	21	AAB56770	Human prostate can
869	5	23.8	316	20	AAY19508	Amino acid sequenc
870	5	23.8	316	22	AAG87803	Mouse T2R09 amino
871	5	23.8	317	22	ABB69511	Drosophila melanog
872	5	23.8	318	21	AAG41938	Arabidopsis thalia
873	5	23.8	318	22	AAR98413	Mouse myeloma MOPC
874	5	23.8	319	17	AAG71491	Mouse myeloma MOPC
875	5	23.8	319	22	AAY49964	Human olfactory re
876	5	23.8	320	21	AAY78311	Human PTB1B mutant
877	5	23.8	322	21	AAG10572	Arabidopsis thalia
878	5	23.8	322	21	AAG32096	Arabidopsis thalia
879	5	23.8	322	22	AAG49691	Arabidopsis thalia
880	5	23.8	323	17	AAR98412	Mouse immunoglobul
881	5	23.8	323	19	AAW48944	Mutant human PTP1B
882	5	23.8	323	19	AAW48936	Mutant human PTP1B
883	5	23.8	323	22	AAU24715	Human olfactory re
884	5	23.8	324	21	AAY49963	Arabidopsis thalia
885	5	23.8	324	21	AAY49963	Newborn mouse immu
886	5	23.8	324	22	AAU52446	Propionibacterium
887	5	23.8	324	22	AAG72697	Murine OR-like pol
888	5	23.8	327	21	AAG10038	Arabidopsis thalia
889	5	23.8	327	22	AAB71709	Carp CTH1 protein.
890	5	23.8	327	22	AAB87795	Mouse T2R01 amino
891	5	23.8	328	20	AAY39999	Human carbonic anh
892	5	23.8	328	20	AAY41737	Human PRO237 prote
893	5	23.8	328	21	AAB44293	Human PRO237 (UNQ2
894	5	23.8	328	21	AAB24041	Human PRO237 prote
895	5	23.8	328	21	AAB19525	Antitumour PRO237
896	5	23.8	328	21	AAY56263	Maedi-Visna virus
897	5	23.8	328	22	AAU12334	Human PRO237 polyp
898	5	23.8	330	22	ABB56472	Murine pCB212 prot
899	5	23.8	330	22	AAB87811	Mouse T2R17 amino
900	5	23.8	331	21	AAY95781	5H7 single chain a
901	5	23.8	331	21	AAU22926	Novel human enzyme
902	5	23.8	332	21	AAG05583	Arabidopsis thalia
903	5	23.8	332	21	AAG06448	Arabidopsis thalia
904	5	23.8	332	21	AAG06484	Arabidopsis thalia
905	5	23.8	332	21	AAG07390	Arabidopsis thalia
906	5	23.8	332	21	AAG24178	Arabidopsis thalia
907	5	23.8	332	21	AAG33349	Zea mays protein f
908	5	23.8	332	21	AAG34599	Arabidopsis thalia
909	5	23.8	332	21	AAG37213	Arabidopsis thalia
910	5	23.8	332	22	AAG37317	Arabidopsis thalia
911	5	23.8	332	22	AAG37989	Arabidopsis thalia
912	5	23.8	332	21	AAG41511	Arabidopsis thalia
913	5	23.8	332	21	AAG44622	Zea mays protein f
914	5	23.8	332	21	AAG44697	Zea mays protein f
915	5	23.8	332	21	AAG47761	Arabidopsis thalia
916	5	23.8	332	21	AAG47770	Arabidopsis thalia
917	5	23.8	332	21	AAG50873	Arabidopsis thalia
918	5	23.8	332	21	AAG50944	Arabidopsis thalia
919	5	23.8	332	21	AAG54243	Arabidopsis thalia
920	5	23.8	333	21	AAB43046	Human ORFX ORF2810
921	5	23.8	333	22	AAM42076	Human polypeptide
922	5	23.8	333	22	AAB82048	Human G protein co
923	5	23.8	335	22	ABB68367	Drosophila melanog
924	5	23.8	335	22	AAU36088	Klebsiella pneumon
925	5	23.8	335	22	AAM93358	Human polypeptide,
926	5	23.8	337	20	AAY00281	Human secreted pro
927	5	23.8	337	22	AAG37208	Arabidopsis thalia
928	5	23.8	337	22	AAU52875	Propionibacterium
929	5	23.8	337	22	AAG67776	Amino acid sequenc
930	5	23.8	339	20	AAY36799	Protein involved i
931	5	23.8	339	21	AAY93406	Amino acid sequenc
932	5	23.8	339	21	ABG26859	Novel human diagno
933	5	23.8	340	11	AAR06830	Thermopsin. Sulph
934	5	23.8	341	9	AAP83200	Sequence encoded b
935	5	23.8	341	22	AAG04645	Arabidopsis thalia
936	5	23.8	341	22	AAM43550	Human polypeptide
937	5	23.8	341	22	AAM43628	Human polypeptide
938	5	23.8	341	22	AAU19933	Novel human calciu
939	5	23.8	342	21	AAG61158	Arabidopsis thalia
940	5	23.8	342	22	AAU01250	B. subtilis acetoh
941	5	23.8	342	22	AAB27240	Human EXMAD-18 SEQ
942	5	23.8	345	21	AAG04644	Arabidopsis thalia
943	5	23.8	345	21	AAG10571	Arabidopsis thalia
944	5	23.8	345	21	AAG47748	Arabidopsis thalia
945	5	23.8	345	22	AAG49690	Human colon cancer
946	5	23.8	345	22	AAG75542	Human colon cancer
947	5	23.8	346	21	AAG61157	Arabidopsis thalia
948	5	23.8	347	14	AAR34224	Amino acid sequenc
949	5	23.8	347	22	AAM39799	Human polypeptide
950	5	23.8	348	20	AAY37541	Protein which is s
951	5	23.8	348	22	AAG71490	Human olfactory re
952	5	23.8	348	22	AAG72907	Human olfactory re
953	5	23.8	348	22	AAE09788	P. pabuli xylogluc
954	5	23.8	354	21	AAG48893	Arabidopsis thalia
955	5	23.8	356	22	AAB32709	Eucalyptus grandis
956	5	23.8	357	22	AAB49469	Canine retrovirus
957	5	23.8	359	21	AAG27740	Arabidopsis thalia
958	5	23.8	360	21	AAG05582	Arabidopsis thalia
959	5	23.8	360	21	AAG06447	Arabidopsis thalia
960	5	23.8	360	21	AAG06483	Arabidopsis thalia

961	5	360	23.8	5	AAG07389	Arabidopsis thalia
962	5	360	23.8	5	AAG24177	Arabidopsis thalia
963	5	360	23.8	5	AAG33348	Zea mays protein f
964	5	360	23.8	5	AAG34598	Arabidopsis thalia
965	5	360	23.8	5	AAG37212	Arabidopsis thalia
966	5	360	23.8	5	AAG37316	Arabidopsis thalia
967	5	360	23.8	5	AAG37988	Arabidopsis thalia
968	5	360	23.8	5	AAG41510	Arabidopsis thalia
969	5	360	23.8	5	AAG44621	Zea mays protein f
970	5	360	23.8	5	AAG44696	Zea mays protein f
971	5	360	23.8	5	AAG47760	Arabidopsis thalia
972	5	360	23.8	5	AAG47769	Arabidopsis thalia
973	5	360	23.8	5	AAG50872	Arabidopsis thalia
974	5	360	23.8	5	AAG50943	Arabidopsis thalia
975	5	360	23.8	5	AAG54242	Arabidopsis thalia
976	5	360	23.8	5	AAG54489	Zea mays protein f
977	5	360	23.8	5	ABB64184	Drosophila melanog
978	5	360	23.8	5	ABB70816	Drosophila melanog
979	5	361	23.8	5	AAR98008	Polypeptide which
980	5	363	23.8	5	AAW24256	Ammonifex histidin
981	5	363	23.8	5	AAW22233	A. thaliana origin
982	5	363	23.8	5	AAW14139	Arabidopsis thalia
983	5	363	23.8	5	AAG05418	Arabidopsis thalia
984	5	363	23.8	5	AAG09923	Arabidopsis thalia
985	5	363	23.8	5	AAG10037	Arabidopsis thalia
986	5	363	23.8	5	AAG43583	Arabidopsis thalia
987	5	363	23.8	5	AAW47771	Arabidopsis thalia
988	5	365	23.8	5	AAW98517	H. pylori GHPO 131
989	5	365	23.8	5	AAG98305	Escherichia coli p
990	5	366	23.8	5	AAW38419	Acyl-coenzyme A:ch
991	5	366	23.8	5	AAG16926	Arabidopsis thalia
992	5	366	23.8	5	AAG47365	Arabidopsis thalia
993	5	367	23.8	5	AAG43582	Arabidopsis thalia
994	5	367	23.8	5	AAU35083	Enterococcus faeca
995	5	368	23.8	5	AAB43084	Human ORFX ORF2848
996	5	368	23.8	5	AAG21032	Arabidopsis thalia
997	5	368	23.8	5	AAG50069	Arabidopsis thalia
998	5	368	23.8	5	AAE01696	Human gene 1 encod
999	5	369	23.8	5	AAU65086	Propionibacterium
1000	5	370	23.8	5	AAY32474	DNA encoding felin

ALIGNMENTS

RESULT 1
ABB70857
ID   ABB70857 standard; Protein; 929 AA.
XX
AC   ABB70857;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 39363.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX

DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14960.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 39363; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   929 AA;

Query Match   33.3%; Score 7; DB 22; Length 929;
Best Local Similarity   100.0%; Pred. No. 58;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0:

Qy   14 SFTQALG 20
               |||||||
Db   103 SFTQALG 109

RESULT 2
AAY40080
ID   AAY40080 standard; Peptide; 27 AA.
XX
AC   AAY40080;
XX
DT   18-NOV-1999 (first entry)
XX
DE   Peptide sequence derived from a human secreted protein.
XX
KW   Secreted protein; gene therapy; cancer; tumor; fetal deficiency;
KW   neurodegenerative disorder; developmental abnormality; blood disorder;
KW   immune system disease; autoimmune disease; leukemia; inflammation;
KW   allergy; Alzheimer's disease; cognitive disorder; schizophrenia;
KW   obesity; osteoporosis; arthritis; infection; AIDS; diabetes; asthma;
KW   connective tissue disorder; transplant rejection; sepsis; acne;
KW   psoriasis; cardiovascular disorder; reproductive disorder;
KW   food additive; food preservative; storage capability.
XX
OS   Homo sapiens.
XX
PN   WO943693-A1.
XX
PD   02-SEP-1999.
XX
PF   24-FEB-1999; 99WO-US03939.
XX
PR   26-FEB-1998; 98US-0076051.
PR   26-FEB-1998; 98US-0076052.
PR   26-FEB-1998; 98US-0076053.
PR   26-FEB-1998; 98US-0076054.
PR   26-FEB-1998; 98US-0076057.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Olsen HS, Florence K, Brewer LA, Ebner R, Ruben SM, Rosen CA;
PI   Duan RD;
XX
DR   WPI; 1999-550857/46.
XX
PT   New human genes and the secreted polypeptides they encode, useful for

Left column:

```
PT    diagnosis and treatment of e.g. cancers, neurological disorders, immune
PT    diseases, inflammation or blood disorders
XX
PS    Disclosure; Page 53; 246pp; English.
XX
CC    AAY4001-92 are derived from human secreted proteins. The
CC    polynucleotides and their corresponding secreted polypeptides are useful
CC    for preventing, treating or ameliorating medical conditions, e.g. by
CC    protein or gene therapy. Pathological conditions can also be diagnosed by
CC    determining the amount of the new polypeptides in a sample or by
CC    determining the presence of mutations in the polynucleotide. Specific
CC    uses include developing products for the diagnosis or treatment of
CC    cancer, tumors, neurodegenerative disorders, developmental abnormalities
CC    and fetal deficiencies, blood disorders, sepsis, diseases of the immune
CC    system, autoimmune diseases, inflammation, allergies, Alzheimer's and
CC    cognitive disorders, schizophrenia, obesity, osteoporosis, arthritis,
CC    infections, AIDS, connective tissue disorders, transplant rejection,
CC    diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders,
CC    and reproductive disorders. The polypeptides or polynucleotides can
CC    also be used as food additives or preservatives, such as to increase
CC    or decrease storage capabilities, fat content, lipid, protein,
CC    carbohydrate, vitamins, minerals, cofactors or other nutritional
CC    components.
XX
SQ    Sequence   27 AA;

Query Match       28.6%;   Score 6;  DB 20;   Length 27;
Best Local Similarity  100.0%;   Pred. No. 26;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   15 FTQALG 20
        ||||||
Db   21 FTQALG 26

RESULT 3
AAM23675
ID    AAM23675 standard; Protein; 48 AA.
XX
AC    AAM23675;
XX
DT    12-OCT-2001  (first entry)
XX
DE    Human EST encoded protein SEQ ID NO: 1200.
XX
KW    Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW    tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW    diagnostics; forensic test; gene mapping; genetic disorder;
KW    biodiversity; gene therapy; nutrition.
XX
OS    Homo sapiens.
XX
PN    WO20154477-A2.
XX
PD    02-AUG-2001.
XX
PF    25-JAN-2001; 2001WO-US02687.
XX
PR    25-JAN-2000; 2000US-0491404.
PR    17-JUL-2000; 2000US-0617746.
PR    03-AUG-2000; 2000US-0631451.
PR    15-SEP-2000; 2000US-0663870.
XX
PA    (HYSE-) HYSEQ INC.
XX
PI    Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI    Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR    WPI; 2001-476164/51.
DR    N-PSDB; AAH98334.
XX
PT    Isolated polypeptide for treatment of diseases, diagnostics, raising
PT    antibodies and research use -
```

Right column:

```
PT    Claim 20; Page 868; 1275pp; English.
PS
XX
CC    The present invention provides the protein and coding sequences of novel
CC    proteins from a variety of organisms, including human, dog, cat, horse,
CC    cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC    urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC    from the organism of interest. They can be used in diagnostics,
CC    forensics, gene mapping, identification of mutations, to assess
CC    biodiversity and for nutritional purposes. The present sequence is a
CC    protein of the invention.
XX
SQ    Sequence   48 AA;

Query Match       28.6%;   Score 6;  DB 22;   Length 48;
Best Local Similarity  100.0%;   Pred. No. 44;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 SGSLST 6
        ||||||
Db   36 SGSLST 41

RESULT 4
AAR94979
ID    AAR94979 standard; Protein; 51 AA.
XX
AC    AAR94979;
XX
DT    05-JAN-1997  (first entry)
XX
DE    Calcium channel isoform CaCh3 IVS3-IVS4 region (ROB3).
XX
KW    Stretch-activated cation channel; SA-Cat; calcium channel; CaCh3;
KW    bone; osteoblast; antisense; osteosclerosis; hypertension.
XX
OS    Rattus sp.
XX
FH    Key            Location/Qualifiers
FT    Domain         1..19
FT                   /note= "transmembrane domain 3"
FT                   /label= IVS3
FT    Domain         33..51
FT                   /note= "transmembrane domain 4"
FT                   /label= IVS4
XX
PN    WO9613269-A1.
XX
PD    09-MAY-1996.
XX
PF    25-OCT-1995;  95WO-US13686.
XX
PR    28-OCT-1994;  94US-0330433.
XX
PA    (DART-) DARTMOUTH COLLEGE.
PA    (JEWI-) JEWISH HOSPITAL ST LOUIS.
XX
PI    Barry ELR,  Duncan RL,  Friedman PA,  Hruska, KA;
XX
DR    WPI; 1996-239267/24.
DR    N-PSDB; AAT18887.
XX
PT    Stretch activated cation channel gene antisense oligonucleotide
PT    used in the treatment of hypertension and osteosclerosis
XX
PS    Example 1; Fig 4; 60pp; English.
XX
CC    Amino acid sequences (AAR94977-79) of UMR-106 clones ROB1, ROB2 and
CC    ROB3 (see also AAT44381 and AAT18886-87) were deduced for the region
CC    including transmembrane domains 3 and 4. ROB1, ROB2 and ROB3 correspond
CC    to isoforms CaCh1, CaCh2 and CaCh3, respectively, of domain 4 of the
CC    alpha 1 subunit of a stretch-activated cation (SA-Cat) channel.
CC    Comparison of the sequences with corresponding sequences of rat.
```

```
CC   brain L-type calcium channels show that CaCh from rat osteosarcoma
CC   UMR-106 cells lack a portion of the IVS3-IV4 linker domain as a
CC   result of alternative splicing.
XX
SQ   Sequence   51 AA;

Query Match             28.6%;  Score 6;  DB 17;  Length 51;
Best Local Similarity  100.0%;  Pred. No. 46;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db    34 TFFRLF 39

RESULT 5
AAR94978
ID   AAR94978 standard; Protein; 51 AA.
XX
AC   AAR94978;
XX
DT   05-JAN-1997  (first entry)
XX
DE   Calcium channel isoform CaCh2 IVS3-IVS4 region (ROB2).
XX
KW   Stretch-activated cation channel; SA-Cat; calcium channel; CaCh2;
KW   bone; osteoblast; antisense; osteosclerosis; hypertension.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..19
FT                   /label= IVS3
FT                   /note= "transmembrane domain 3"
FT   Domain          33..51
FT                   /label= IVS4
FT                   /note= "transmembrane domain 4"
XX
PN   WO9613269-A1.
XX
PD   09-MAY-1996.
XX
PF   25-OCT-1995;  95WO-US13686.
XX
PR   28-OCT-1994;  94US-0330433.
XX
PA   (DART-) DARTMOUTH COLLEGE.
PA   (JEWI-) JEWISH HOSPITAL ST LOUIS.
XX
PI   Barry ELR,  Duncan RL,  Friedman PA,  Hruska KA;
XX
DR   WPI; 1996-239267/24.
DR   N-PSDB; AAT18886.
XX
PT   Stretch activated cation channel gene antisense oligonucleotide
PT   used in the treatment of hypertension and osteosclerosis
XX
PS   Example 1; Fig 4; 60pp; English.
XX
CC   Amino acid sequences (AAR94977-79) of UMR-106 clones ROB1, ROB2 and
CC   ROB3 (see also AAT44381 and AAT18886-87) were deduced for the region
CC   including transmembrane domains 3 and 4.  ROB1, ROB2 and ROB3 correspond
CC   to isoforms CaCh1, CaCh2 and CaCh3, respectively, of domain 4 of the
CC   alpha 1 subunit of a stretch-activated cation (SA-Cat) channel.
CC   Comparison of the sequences with corresponding sequences of rat
CC   brain L-type calcium channels show that CaCh from rat osteosarcoma
CC   UMR-106 cells lack a portion of the IVS3-IV4 linker domain as a
CC   result of alternative splicing.  ROB2 is a close isoform of the
CC   SA-Cat channel.
XX
SQ   Sequence   51 AA;

Query Match             28.6%;  Score 6;  DB 17;  Length 51;
```

```
Best Local Similarity  100.0%;  Pred. No. 46;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db    34 TFFRLF 39

RESULT 6
AAR97851
ID   AAR97851 standard; Protein; 62 AA.
XX
AC   AAR97851;
XX
DT   05-JAN-1997  (first entry)
XX
DE   Rat brain calcium channel IVS3-IVS4 region (RatBr2).
XX
KW   Stretch-activated cation channel; SA-Cat; calcium channel; CaCh;
KW   bone; osteoblast; antisense; osteosclerosis; hypertension.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..19
FT                   /label= IVS3
FT                   /note= "transmembrane domain 3"
FT   Domain          44..62
FT                   /label= IVS4
FT                   /note= "transmembrane domain 4"
XX
PN   WO9613269-A1.
XX
PD   09-MAY-1996.
XX
PF   25-OCT-1995;  95WO-US13686.
XX
PR   28-OCT-1994;  94US-0330433.
XX
PA   (DART-) DARTMOUTH COLLEGE.
PA   (JEWI-) JEWISH HOSPITAL ST LOUIS.
XX
PI   Barry ELR,  Duncan RL,  Friedman PA,  Hruska KA;
XX
DR   WPI; 1996-239267/24.
XX
PT   Stretch activated cation channel gene antisense oligonucleotide
PT   used in the treatment of hypertension and osteosclerosis
XX
PS   Example 1; Fig 4; 60pp; English.
XX
CC   Amino acid sequences (AAR97850-52) for the transmembrane domain
CC   3 to 4 region of rat brain calcium channel proteins RabSkel1,
CC   RatBr2 and RatBr3, respectively, were compared to corresponding
CC   regions of calcium channel CaCh1 (ROB1), CaCh2 (ROB2) and CaCh3
CC   (ROB3) isoforms (see also AAR94977-79) deduced from rat osteosarcoma
CC   UMR-106 cDNA clones (AAT44381 and AAT18886-87).  The comparison revealed
CC   the use of an alternative splice acceptor site or an exon skipping
CC   event in the IVS3-IVS4 linker, producing shorter transcripts in ROB1-3.
XX
SQ   Sequence   62 AA;

Query Match             28.6%;   Score 6;  DB 17;  Length 62;
Best Local Similarity  100.0%;   Pred. No. 55;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db    45 TFFRLF 50

RESULT 7
AAR97852
```

```
ID   AAR97852 standard; Protein; 66 AA.
XX
AC   AAR97852;
XX
DT   05-JAN-1997  (first entry)
XX
DE   Rat brain calcium channel IVS3-IVS4 region (RatBr3).
XX
KW   Stretch-activated cation channel; SA-Cat; calcium channel; CaCh;
KW   bone; osteoblast; antisense; osteosclerosis; hypertension.
XX
OS   Rattus sp.
XX
FH   Key            Location/Qualifiers
FT   Domain         1..19
FT                  /label= IVS3
FT                  /note= "transmembrane domain 3"
FT   Domain         48..66
FT                  /label= IVS4
FT                  /note= "transmembrane domain 4"
XX
PN   WO9613269-A1.
XX
PD   09-MAY-1996.
XX
PF   25-OCT-1995;  95WO-US13686.
XX
PR   28-OCT-1994;  94US-0330433.
XX
PA   (DART-) DARTMOUTH COLLEGE.
PA   (JEWI-) JEWISH HOSPITAL ST LOUIS.
XX
PI   Barry ELR, Duncan RL, Friedman PA, Hruska KA;
XX
DR   WPI; 1996-239267/24.
XX
PT   Stretch activated cation channel gene antisense oligonucleotide -
PT   used in the treatment of hypertension and osteosclerosis
XX
PS   Example 1; Fig 4; 60pp; English.
XX
CC   Amino acid sequences (AAR97850-52) for the transmembrane domain
CC   3 to 4 region of rat brain calcium channel proteins RabSkel1,
CC   RatBr2 and RatBr3, respectively, were compared to corresponding
CC   regions of calcium channel CaCh1 (ROB1), CaCh2 (ROB2) and CaCh3
CC   (ROB3) isoforms (see also AAR94977-79) deduced from rat osteosarcoma
CC   UMR-106 cDNA clones (AAT44381 and AAT18886-87).  The comparison revealed
CC   the use of an alternative splice acceptor site or an exon skipping
CC   event in the IVS3-IVS4 linker, producing shorter transcripts in ROB1-3.
XX
SQ   Sequence  66 AA;

Query Match             28.6%;  Score 6;  DB 17;  Length 66;
Best Local Similarity   100.0%;  Pred. No. 58;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 TFPRLF 11
          ||||||
Db     49 TFPRLF 54


RESULT 8
ABG04703
ID   ABG04703 standard; Protein; 66 AA.
XX
AC   ABG04703;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #4694.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS68890.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 35062; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  66 AA;

Query Match             28.6%;  Score 6;  DB 22;  Length 66;
Best Local Similarity   100.0%;  Pred. No. 58;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GSLSTF 7
          ||||||
Db     28 GSLSTF 33


RESULT 9
AAM89377
ID   AAM89377 standard; Protein; 78 AA.
XX
AC   AAM89377;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen SEQ ID NO:16970.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   2001WO-US01354.
XX
PR   17-JAN-2001;  2001WO-US01354.
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0186350.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   11-JUL-2000;  2000US-0217496.
PR   14-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225757.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
PR   30-AUG-2000;  2000US-0228924.
PR   01-SEP-2000;  2000US-0229287.
PR   01-SEP-2000;  2000US-0229343.
PR   01-SEP-2000;  2000US-0229344.
PR   01-SEP-2000;  2000US-0229345.
PR   05-SEP-2000;  2000US-0229509.
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.

PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   13-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   05-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251858.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-)  HUMAN GENOME SCI INC.
XX
```

PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR N-PSDB; AAK62158.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 11; SEQ ID NO 16970; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 78 AA;

Query Match 28.6%; Score 6; DB 22; Length 78;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GSLSTF 7
       ||||||
Db 22 GSLSTF 27

RESULT 10
AAW59796
ID AAW59796 standard; Protein; 96 AA.
XX
AC AAW59796;
XX
DT 12-OCT-1998 (first entry)
XX
DE Nucleotide sequence of complete EST3 gene product 2.
XX
KW EST3 gene; telomerase inhibitor; cancer; mammal; tumour growth;
KW stem cell; chemotherapeutic agent.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9823759-A2.
XX
PD 04-JUN-1998.
XX
PF 26-NOV-1997; 97WO-US21272.
XX
PR 26-NOV-1996; 96US-0756693.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Lundblad V;
XX
DR WPI; 1998-332927/29.
DR N-PSDB; AAV41581.
XX
PT Telomerase-associated proteins useful to identify telomerase
PT inhibitors - useful e.g. for cancer treatment in mammals, and
PT screening methods to isolate additional telomerase-associated genes

PT or protein homologues
XX
PS Disclosure; Fig 7; 71pp; English.
XX
CC This is the nucleotide sequence of the EST3 gene product 2 (including
CC flanking regions and the +1 ribosomal frame shift), used in the method
CC of the invention to identify telomerase inhibitors which are useful in
CC the treatment of cancer. The proteins are useful to isolate
CC telomerase-inhibiting compounds. Such inhibitors are useful in cancer
CC treatment in mammals, since reactivation of telomerase (normally present
CC only in germ line cells) is thought to be necessary for sustained tumour
CC growth, and only tumour and stem cells would be targeted by such agents,
CC producing limited side effects compared to chemotherapeutic agents.
XX
SQ Sequence 96 AA;

Query Match 28.6%; Score 6; DB 19; Length 96;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GSLSTF 7
       ||||||
Db 84 GSLSTF 89

RESULT 11
AAY40074
ID AAY40074 standard; Peptide; 122 AA.
XX
AC AAY40074;
XX
DT 18-NOV-1999 (first entry)
XX
DE Peptide sequence derived from a human secreted protein.
XX
KW Secreted protein; gene therapy; cancer; tumor; fetal deficiency;
KW neurodegenerative disorder; developmental abnormality; blood disorder;
KW immune system disease; autoimmune disease; leukemia; inflammation;
KW allergy; Alzheimer's disease; cognitive disorder; schizophrenia;
KW obesity; osteoporosis; arthritis; infection; AIDS; diabetes; asthma;
KW connective tissue disorder; transplant rejection; sepsis; acne;
KW psoriasis; cardiovascular disorder; reproductive disorder;
KW food additive; food preservative; storage capability.
XX
OS Homo sapiens.
XX
PN WO9943693-A1.
XX
PD 02-SEP-1999.
XX
PF 24-FEB-1999; 99WO-US03939.
XX
PR 26-FEB-1998; 98US-0076051.
PR 26-FEB-1998; 98US-0076052.
PR 26-FEB-1998; 98US-0076053.
PR 26-FEB-1998; 98US-0076054.
PR 26-FEB-1998; 98US-0076057.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Olsen HS, Florence K, Brewer LA, Ebner R, Ruben SM, Rosen CA;
PI Duan RD;
XX
DR WPI; 1999-550857/46.
XX
PT New human genes and the secreted polypeptides they encode, useful for
PT diagnosis and treatment of e.g. cancers, neurological disorders, immune
PT diseases, inflammation or blood disorders
XX
PS Disclosure; Page 53; 246pp; English.
XX
CC AAY40001-92 are derived from human secreted proteins. The
CC polynucleotides and their corresponding secreted polypeptides are useful

```
CC   for preventing, treating or ameliorating medical conditions, e.g. by
CC   protein or gene therapy. Pathological conditions can also be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the polynucleotide. Specific
CC   uses include developing products for the diagnosis or treatment of
CC   cancer, tumors, neurodegenerative disorders, developmental abnormalities
CC   and fetal deficiencies, blood disorders, sepsis, diseases of the immune
CC   system, autoimmune diseases, inflammation, allergies, Alzheimer's and
CC   cognitive disorders, schizophrenia, obesity, osteoporosis, arthritis,
CC   infections, AIDS, connective tissue disorders, transplant rejection,
CC   diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders,
CC   and reproductive disorders. The polypeptides or polynucleotides can
CC   also be used as food additives or preservatives, such as to increase
CC   or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors or other nutritional
CC   components.
XX
SQ   Sequence   122 AA;

Query Match             28.6%;  Score 6;  DB 20;  Length 122;
Best Local Similarity   100.0%;  Pred. No. 99;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15  FTQALG 20
               ||||||
Db         97  FTQALG 102

RESULT 12
AAU45360
ID   AAU45360 standard; Protein; 141 AA.
XX
AC   AAU45360;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #6256.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI   L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59525.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 6555; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   141 AA;

Query Match             28.6%;  Score 6;  DB 22;  Length 141;
Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         14  SFTQAL 19
               ||||||
Db         47  SFTQAL 52

RESULT 13
AAE11946
ID   AAE11946 standard; Protein; 147 AA.
XX
AC   AAE11946;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human CCAAT/enhancer binding protein (C/EBP) beta-3 or p20 isoform.
XX
KW   Human; CCAAT/enhancer binding protein; C/EBPbeta; transcription factor;
KW   interleukin; IL; p20; inflammation; adult respiratory distress syndrome;
KW   allergic rhinitis; arthritis; bronchitis; bronchopulmonary dysplasia;
KW   cystic fibrosis; extensive allergic alveolitis; anti-inflammatory;
KW   idiopathic pulmonary fibrosis; interstitial lung disease; anti-allergic;
KW   inflammatory bowel disease; respiratory viral infection; anti-arthritic;
KW   anti-asthma; intestinal; antiviral.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         66..82
FT                   /label= NLS_A
FT                   /note= "Nuclear localisation sequence A"
FT   Modified-site   68
FT                   /note= "Phosphorylation site"
FT   Peptide         79..95
FT                   /label= NLS_B
FT                   /note= "Nuclear localisation sequence B"
XX
PN   WO200160320-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05578.
XX
PR   18-FEB-2000; 2000US-0183584.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Bringham KL, Stecenko AA, Sealy L;
XX
DR   WPI; 2001-581897/65.
DR   N-PSDB; AAD19382.
XX
```

```
PT   Treating inflammation, particularly of the lung, by increasing activity
PT   of p20, the beta3-isoform of CCAAT/enhancer binding protein -
XX
PS   Example 5; Fig 5; 200pp; English.
XX
CC   The present sequence is a human CCAAT/Enhancer Binding Protein (C/EBP)
CC   beta-3 (referred as p20) isoform. The C/EBPbeta is a transcription factor
CC   which is identified as being critical for maximal interleukin (iL)-6 and
CC   IL-8 expression. The isoforms of C/EBPbeta are C/EBPbeta-1, C/EBPbeta-2
CC   and C/EBPbeta-3. The p20 isoform of C/EBPbeta is useful for treating
CC   inflammation, adult respiratory distress syndrome, allergic rhinitis,
CC   arthritis, bronchitis, bronchopulmonary dysplasia, cystic fibrosis,
CC   extensive allergic alveolitis, idiopathic pulmonary fibrosis,
CC   inflammatory bowel disease, interstitial lung disease and respiratory
CC   viral infection.
XX
SQ   Sequence   147 AA;

Query Match             28.6%;  Score 6;  DB 22;  Length 147;
Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1  SGSLST  6
          ||||||
Db    23  SGSLST  28


RESULT 14
AAU16099
ID   AAU16099 standard; Protein; 150 AA.
XX
AC   AAU16099;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1052.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
```

```
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   05-DEC-2000;  2000US-0256719.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2001US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26086.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 11; SEQ ID No 1052; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC   are also used in diagnosing a pathological condition or susceptibility
CC   to a pathological condition. Antibodies to the proteins can also
CC   be used in alleviating symptoms associated with the disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC   and ocular disorders e.g. corneal infection, and many other
CC   disorders listed in the specification. The polypeptides can also
CC   be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention.

Query Match             28.6%;  Score 6;  DB 22;  Length 150;
Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    13 RSFTQA 18
         ||||||
Db   119 RSFTQA 124

RESULT 15
AAG48562
ID   AAG48562 standard; Protein; 152 AA.
XX
AC   AAG48562;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 61338.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   06-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131149.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
```

us-09-833-017b-4.oligo.rag

```
PR  24-MAY-1999; 99US-0135629.        PR  02-AUG-1999; 99US-0146389.
PR  25-MAY-1999; 99US-0136021.        PR  03-AUG-1999; 99US-0147038.
PR  27-MAY-1999; 99US-0136392.        PR  04-AUG-1999; 99US-0147204.
PR  28-MAY-1999; 99US-0136782.        PR  04-AUG-1999; 99US-0147302.
PR  01-JUN-1999; 99US-0137222.        PR  05-AUG-1999; 99US-0147192.
PR  03-JUN-1999; 99US-0137528.        PR  05-AUG-1999; 99US-0147260.
PR  04-JUN-1999; 99US-0137502.        PR  06-AUG-1999; 99US-0147303.
PR  07-JUN-1999; 99US-0137724.        PR  06-AUG-1999; 99US-0147416.
PR  08-JUN-1999; 99US-0138094.        PR  09-AUG-1999; 99US-0147493.
PR  10-JUN-1999; 99US-0138540.        PR  10-AUG-1999; 99US-0147935.
PR  10-JUN-1999; 99US-0138847.        PR  10-AUG-1999; 99US-0148171.
PR  14-JUN-1999; 99US-0139119.        PR  11-AUG-1999; 99US-0148319.
PR  16-JUN-1999; 99US-0139452.        PR  12-AUG-1999; 99US-0148341.
PR  16-JUN-1999; 99US-0139453.        PR  13-AUG-1999; 99US-0148565.
PR  17-JUN-1999; 99US-0139492.        PR  13-AUG-1999; 99US-0148684.
PR  18-JUN-1999; 99US-0139454.        PR  16-AUG-1999; 99US-0149368.
PR  18-JUN-1999; 99US-0139455.        PR  17-AUG-1999; 99US-0149175.
PR  18-JUN-1999; 99US-0139457.        PR  18-AUG-1999; 99US-0149426.
PR  18-JUN-1999; 99US-0139458.        PR  20-AUG-1999; 99US-0149722.
PR  18-JUN-1999; 99US-0139459.        PR  20-AUG-1999; 99US-0149723.
PR  18-JUN-1999; 99US-0139460.        PR  20-AUG-1999; 99US-0149929.
PR  18-JUN-1999; 99US-0139461.        PR  23-AUG-1999; 99US-0149902.
PR  18-JUN-1999; 99US-0139462.        PR  23-AUG-1999; 99US-0149930.
PR  18-JUN-1999; 99US-0139463.        PR  25-AUG-1999; 99US-0150566.
PR  18-JUN-1999; 99US-0139750.        PR  26-AUG-1999; 99US-0150884.
PR  18-JUN-1999; 99US-0139763.        PR  27-AUG-1999; 99US-0151065.
PR  21-JUN-1999; 99US-0139817.        PR  27-AUG-1999; 99US-0151066.
PR  22-JUN-1999; 99US-0139899.        PR  27-AUG-1999; 99US-0151080.
PR  23-JUN-1999; 99US-0140353.        PR  30-AUG-1999; 99US-0151303.
PR  24-JUN-1999; 99US-0140354.        PR  31-AUG-1999; 99US-0151438.
PR  28-JUN-1999; 99US-0140695.        PR  01-SEP-1999; 99US-0151930.
PR  29-JUN-1999; 99US-0140823.        PR  07-SEP-1999; 99US-0152363.
PR  30-JUN-1999; 99US-0140991.        PR  10-SEP-1999; 99US-0153070.
PR  01-JUL-1999; 99US-0141842.        PR  13-SEP-1999; 99US-0153758.
PR  01-JUL-1999; 99US-0142154.        PR  15-SEP-1999; 99US-0154018.
PR  02-JUL-1999; 99US-0142055.        PR  16-SEP-1999; 99US-0154039.
PR  06-JUL-1999; 99US-0142390.        PR  20-SEP-1999; 99US-0154779.
PR  08-JUL-1999; 99US-0142803.        PR  22-SEP-1999; 99US-0155139.
PR  09-JUL-1999; 99US-0142920.        PR  23-SEP-1999; 99US-0155486.
PR  12-JUL-1999; 99US-0142977.        PR  24-SEP-1999; 99US-0155659.
PR  13-JUL-1999; 99US-0143542.        PR  28-SEP-1999; 99US-0156458.
PR  14-JUL-1999; 99US-0143624.        PR  29-SEP-1999; 99US-0156596.
PR  15-JUL-1999; 99US-0144005.        PR  04-OCT-1999; 99US-0157117.
PR  16-JUL-1999; 99US-0144085.        PR  05-OCT-1999; 99US-0157753.
PR  16-JUL-1999; 99US-0144086.        PR  06-OCT-1999; 99US-0157865.
PR  19-JUL-1999; 99US-0144325.        PR  07-OCT-1999; 99US-0158029.
PR  19-JUL-1999; 99US-0144331.        PR  08-OCT-1999; 99US-0158232.
PR  19-JUL-1999; 99US-0144332.        PR  12-OCT-1999; 99US-0158369.
PR  19-JUL-1999; 99US-0144333.        PR  13-OCT-1999; 99US-0159293.
PR  19-JUL-1999; 99US-0144334.        PR  13-OCT-1999; 99US-0159294.
PR  19-JUL-1999; 99US-0144335.        PR  14-OCT-1999; 99US-0159295.
PR  20-JUL-1999; 99US-0144352.        PR  14-OCT-1999; 99US-0159929.
PR  20-JUL-1999; 99US-0144632.        PR  14-OCT-1999; 99US-0159330.
PR  20-JUL-1999; 99US-0144884.        PR  14-OCT-1999; 99US-0159331.
PR  21-JUL-1999; 99US-0144814.        PR  18-OCT-1999; 99US-0159637.
PR  21-JUL-1999; 99US-0145086.        PR  21-OCT-1999; 99US-0159638.
PR  21-JUL-1999; 99US-0145088.        PR  21-OCT-1999; 99US-0159584.
PR  22-JUL-1999; 99US-0145085.        PR  21-OCT-1999; 99US-0160741.
PR  22-JUL-1999; 99US-0145087.        PR  21-OCT-1999; 99US-0160767.
PR  22-JUL-1999; 99US-0145089.        PR  21-OCT-1999; 99US-0160768.
PR  22-JUL-1999; 99US-0145192.        PR  21-OCT-1999; 99US-0160770.
PR  23-JUL-1999; 99US-0145145.        PR  22-OCT-1999; 99US-0160814.
PR  23-JUL-1999; 99US-0145218.        PR  22-OCT-1999; 99US-0160815.
PR  23-JUL-1999; 99US-0145224.        PR  22-OCT-1999; 99US-0160980.
PR  26-JUL-1999; 99US-0145276.        PR  25-OCT-1999; 99US-0160981.
PR  27-JUL-1999; 99US-0145913.        PR  25-OCT-1999; 99US-0160989.
PR  27-JUL-1999; 99US-0145918.        PR  26-OCT-1999; 99US-0161404.
PR  27-JUL-1999; 99US-0145919.        PR  26-OCT-1999; 99US-0161405.
PR  28-JUL-1999; 99US-0145951.        PR  26-OCT-1999; 99US-0161406.
PR  02-AUG-1999; 99US-0146386.        PR  26-OCT-1999; 99US-0161359.
PR  02-AUG-1999; 99US-0146388.        PR  26-OCT-1999; 99US-0161360.
                                      PR  26-OCT-1999; 99US-0161361.
                                      PR  28-OCT-1999; 99US-0161920.
```

PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   28.6%;   Score 6; DB 21; Length 152;
Best Local Similarity   100.0%; Pred. No.1.2e+02;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
          |||||||
Db   84 SGSLST 89

Search completed: November 5, 2002, 11:05:26
Job time : 44.6119 secs

```
           GenCore version 5.1.3
      Copyright (c) 1993 - 2002 Compugen Ltd.
```

OM protein - protein search, using sw model

Run on: November 5, 2002, 11:03:27 ; Search time 6.89552 Seconds
                                         (without alignments)
                                         74.387 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 21
Sequence: 1 SGSLSTFRFLFNRSFTQALGK 21

Scoring table: OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size :   0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   Issued_Patents_AA:*
         1:   /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
         2:   /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
         3:   /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
         4:   /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
         5:   /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
         6:   /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	28.6	282	4	US-08-482-918-56	Sequence 56, Appl
2	6	28.6	282	4	US-09-224-681-56	Sequence 56, Appl
3	6	28.6	282	4	US-08-336-728A-56	Sequence 56, Appl
4	6	28.6	446	4	US-09-268-992-51	Sequence 51, Appl
5	6	28.6	466	4	US-09-268-992-64	Sequence 64, Appl
6	6	28.6	477	4	US-09-268-992-4	Sequence 4, Appl
7	6	28.6	495	2	US-09-268-992-2	Sequence 2, Appl
8	6	28.6	499	2	US-09-952-853-6	Sequence 6, Appl
9	6	28.6	499	2	US-08-914-848-6	Sequence 6, Appl
10	6	28.6	635	2	US-09-014-969-11	Sequence 11, Appl
11	6	28.6	1968	1	US-07-745-206A-7	Sequence 7, Appl
12	6	28.6	1968	2	US-08-455-543A-45	Sequence 45, Appl
13	6	28.6	1968	2	US-08-223-305C-45	Sequence 45, Appl
14	6	28.6	1968	2	US-08-311-363-7	Sequence 7, Appl
15	6	28.6	2161	1	US-07-745-206A-2	Sequence 2, Appl
16	6	28.6	2161	1	US-08-455-543A-49	Sequence 49, Appl
17	6	28.6	2161	1	US-08-455-543A-51	Sequence 51, Appl
18	6	28.6	2161	2	US-08-223-305C-49	Sequence 49, Appl
19	6	28.6	2161	2	US-08-223-305C-51	Sequence 51, Appl
20	6	28.6	2161	2	US-08-311-363-2	Sequence 2, Appl
21	5	23.8	8	4	US-09-139-802-4	Sequence 4, Appl
22	5	23.8	15	2	US-08-726-464B-30	Sequence 30, Appl
23	5	23.8	16	2	US-08-480-190-261	Sequence 261, App
24	5	23.8	16	2	US-08-488-379-261	Sequence 261, App
25	5	23.8	16	5	PCT-US93-07545-261	Sequence 261, App
26	5	23.8	23	3	US-08-383-753-82	Sequence 82, Appl
27	5	23.8	23	2	US-08-586-772-82	Sequence 82, Appl
28	5	23.8	23	2	US-08-959-512-82	Sequence 82, Appl
29	5	23.8	23	4	US-09-512-983-82	Sequence 82, Appl
30	5	23.8	31	4	US-09-177-249-21	Sequence 21, Appl
31	5	23.8	73	2	US-08-245-511-36	Sequence 36, Appl
32	5	23.8	73	2	US-08-600-993A-36	Sequence 36, Appl
33	5	23.8	75	4	US-09-042-012-8	Sequence 8, Appli
34	5	23.8	75	4	US-09-305-086-2	Sequence 2, Appli
35	5	23.8	75	1	US-09-457-324-8	Sequence 8, Appli
36	5	23.8	78	4	US-08-680-726A-60	Sequence 60, Appl
37	5	23.8	78	4	US-09-092-409-60	Sequence 60, Appl
38	5	23.8	97	3	US-08-300-928C-10	Sequence 10, Appl
39	5	23.8	97	3	US-08-430-944D-10	Sequence 10, Appl
40	5	23.8	97	3	US-08-430-014-10	Sequence 10, Appl
41	5	23.8	97	3	US-08-431-184-10	Sequence 10, Appl
42	5	23.8	97	3	US-08-858-207A-431	Sequence 431, App
43	5	23.8	109	1	US-07-662-193-5	Sequence 5, Appli
44	5	23.8	109	1	US-07-807-529A-6	Sequence 6, Appli
45	5	23.8	109	3	US-08-300-928C-8	Sequence 8, Appli
46	5	23.8	109	3	US-08-430-944D-8	Sequence 8, Appli
47	5	23.8	109	3	US-08-430-014-8	Sequence 8, Appli
48	5	23.8	109	3	US-08-431-184-8	Sequence 8, Appli
49	5	23.8	109	5	PCT-US93-02462-6	Sequence 6, Appli
50	5	23.8	111	4	US-07-662-193-4	Sequence 4, Appli
51	5	23.8	111	3	US-08-300-928C-6	Sequence 6, Appli
52	5	23.8	111	3	US-08-430-944D-6	Sequence 6, Appli
53	5	23.8	111	3	US-08-430-014-6	Sequence 6, Appli
54	5	23.8	111	3	US-08-431-184-6	Sequence 6, Appli
55	5	23.8	116	1	US-08-139-862-4	Sequence 4, Appli
56	5	23.8	117	4	US-09-046-479-2	Sequence 2, Appli
57	5	23.8	120	4	US-08-446-345-27	Sequence 27, Appl
58	5	23.8	123	1	US-08-082-623-3	Sequence 3, Appli
59	5	23.8	125	4	US-08-751-359-13	Sequence 13, Appl
60	5	23.8	125	4	US-08-907-146-13	Sequence 13, Appl
61	5	23.8	126	4	US-08-751-359-12	Sequence 12, Appl
62	5	23.8	126	4	US-08-907-146-12	Sequence 12, Appl
63	5	23.8	130	3	US-09-014-574-1	Sequence 1, Appli
64	5	23.8	156	4	US-09-181-183-2	Sequence 2, Appli
65	5	23.8	156	4	US-09-277-700-2	Sequence 2, Appli
66	5	23.8	169	4	US-08-720-625-4	Sequence 4, Appli
67	5	23.8	170	2	US-08-764-563-5	Sequence 5, Appli
68	5	23.8	172	3	US-08-557-309B-39	Sequence 39, Appl
69	5	23.8	172	3	US-08-834-306-39	Sequence 39, Appl
70	5	23.8	172	4	US-08-993-674A-39	Sequence 39, Appl
71	5	23.8	179	3	US-08-764-563-4	Sequence 4, Appli
72	5	23.8	179	4	US-08-862-124-2	Sequence 2, Appli
73	5	23.8	196	4	US-08-685-992-1	Sequence 1, Appli
74	5	23.8	196	2	US-09-144-925-1	Sequence 1, Appli
75	5	23.8	196	2	US-08-484-126-7	Sequence 7, Appli
76	5	23.8	200	6	5189147-8	Patent No. 5189147
77	5	23.8	212	4	US-08-737-129A-2	Sequence 2, Appli
78	5	23.8	217	5	PCT-US94-14106-59	Sequence 59, Appl
79	5	23.8	219	1	US-08-152-019A-31	Sequence 31, Appl
80	5	23.8	219	2	US-08-152-019A-32	Sequence 32, Appl
81	5	23.8	219	2	US-08-460-309-18	Sequence 18, Appl
82	5	23.8	219	2	US-08-125-077-18	Sequence 18, Appl
83	5	23.8	222	4	US-09-052-089A-3	Sequence 3, Appli
84	5	23.8	222	2	US-08-737-129A-6	Sequence 6, Appli
85	5	23.8	222	4	US-09-181-183-28	Sequence 28, Appl
86	5	23.8	222	4	US-09-277-700-28	Sequence 28, Appl
87	5	23.8	223	5	PCT-US94-14106-51	Sequence 51, Appl
88	5	23.8	236	5	PCT-US94-14106-55	Sequence 55, Appl
89	5	23.8	236	1	US-08-158-682A-4	Sequence 4, Appli
90	5	23.8	236	1	US-08-687-895-4	Sequence 4, Appli
91	5	23.8	236	2	US-08-816-241-4	Sequence 4, Appli
92	5	23.8	236	2	US-09-040-482-4	Sequence 4, Appli
93	5	23.8	239	3	US-09-128-395-4	Sequence 4, Appli
94	5	23.8	239	3	US-08-916-902A-1	Sequence 1, Appli
95	5	23.8	239	1	US-09-213-389-1	Sequence 1, Appli
96	5	23.8	240	3	US-09-194-905-8	Sequence 8, Appli
97	5	23.8	243	4	US-09-286-690-10	Sequence 10, Appl
98	5	23.8	247	5	US-09-100-804-13	Sequence 13, Appl
99	5	23.8	247	2	PCT-US94-07659-2	Sequence 2, Appli
100	5	23.8	254	2	US-08-792-824-4	Sequence 4, Appli

#			Len	Cnt	Patent ID	Description
101	23.8	5	254	2	US-08-792-824-7	Sequence 7, Appli
102	23.8	5	254	2	US-08-792-824-10	Sequence 10, Appli
103	23.8	5	254	2	US-08-792-824-13	Sequence 13, Appli
104	23.8	5	269	1	US-08-203-716-3	Sequence 3, Appli
105	23.8	5	269	2	US-08-440-179-3	Sequence 3, Appli
106	23.8	5	269	1	US-08-432-693-11	Sequence 11, Appli
107	23.8	5	269	2	US-08-599-695-11	Sequence 11, Appli
108	23.8	5	269	2	US-08-483-806-2	Sequence 2, Appli
109	23.8	5	269	3	US-09-211-290-11	Sequence 11, Appli
110	23.8	5	269	3	US-09-030-613-11	Sequence 11, Appli
111	23.8	5	269	3	US-09-322-676-11	Sequence 11, Appli
112	23.8	5	269	4	US-09-039-657-3	Sequence 3, Appli
113	23.8	5	269	4	US-08-748-547-4	Sequence 4, Appli
114	23.8	5	269	4	US-08-466-036A-11	Sequence 11, Appli
115	23.8	5	269	4	US-09-451-905-11	Sequence 11, Appli
116	23.8	5	269	5	PCT-US91-02339-1	Sequence 1, Appli
117	23.8	5	282	5	PCT-US94-09752-3	Sequence 3, Appli
118	23.8	5	287	1	US-08-457-245-3	Sequence 3, Appli
119	23.8	5	287	3	US-08-862-124-17	Sequence 17, Appli
120	23.8	5	288	3	US-09-100-604-14	Sequence 14, Appli
121	23.8	5	297	3	US-08-534-910B-7	Sequence 7, Appli
123	23.8	5	297	4	US-08-534-910B-10	Sequence 10, Appli
124	23.8	5	304	2	US-08-862-124-14	Sequence 14, Appli
125	23.8	5	320	3	US-08-579-940-8	Sequence 8, Appli
126	23.8	5	320	3	US-08-964-208-2	Sequence 2, Appli
127	23.8	5	320	3	US-08-964-313-2	Sequence 2, Appli
128	23.8	5	320	4	US-08-069-138-2	Sequence 2, Appli
129	23.8	5	322	3	US-08-964-308-8	Sequence 8, Appli
130	23.8	5	322	3	US-08-964-313-8	Sequence 8, Appli
131	23.8	5	324	4	US-09-069-138-8	Sequence 8, Appli
132	23.8	5	328	2	US-08-579-940-7	Sequence 7, Appli
133	23.8	5	345	2	US-08-446-445-40	Sequence 40, Appli
134	23.8	5	361	1	US-08-537-434-1	Sequence 1, Appli
135	23.8	5	363	1	US-08-484-105-20	Sequence 20, Appli
136	23.8	5	363	1	US-08-484-106-20	Sequence 20, Appli
137	23.8	5	363	4	US-08-646-590B-36	Sequence 36, Appli
138	23.8	5	363	4	US-08-412-184-36	Sequence 36, Appli
139	23.8	5	369	1	US-08-700-359-21	Sequence 21, Appli
140	23.8	5	370	3	US-08-911-321-10	Sequence 10, Appli
141	23.8	5	374	3	US-08-609-236-6	Sequence 6, Appli
142	23.8	5	375	2	US-08-494-151-14	Sequence 14, Appli
143	23.8	5	375	4	US-09-106-217-16	Sequence 16, Appli
144	23.8	5	375	4	US-09-171-337A-16	Sequence 16, Appli
145	23.8	5	376	1	US-09-171-337A-8	Sequence 8, Appli
146	23.8	5	377	2	US-08-588-113-2	Sequence 2, Appli
147	23.8	5	379	3	US-09-106-217-2	Sequence 2, Appli
148	23.8	5	379	4	US-08-840-204-3	Sequence 3, Appli
149	23.8	5	385	1	US-09-193-266-1	Sequence 1, Appli
150	23.8	5	385	1	US-09-144-914-4	Sequence 5, Appli
151	23.8	5	390	1	US-08-036-210-5	Sequence 5, Appli
152	23.8	5	390	2	US-08-449-609-5	Sequence 6, Appli
153	23.8	5	390	5	US-08-121-714-6	Sequence 6, Appli
154	23.8	5	392	2	US-08-477-108A-6	Sequence 1, Appli
155	23.8	5	392	4	US-08-315-461-7	Sequence 2, Appli
156	23.8	5	394	4	US-08-477-112-6	Sequence 7, Appli
157	23.8	5	397	2	PCT-US93-08322-6	Sequence 2, Appli
158	23.8	5	398	1	US-08-886-152-1	Sequence 3, Appli
159	23.8	5	402	3	US-08-886-152-3	Sequence 12, Appli
160	23.8	5	402	3	US-09-079-415-2	Sequence 4, Appli
161	23.8	5	402	3	US-08-328-314-2	Sequence 2, Appli
162	23.8	5	412	4	US-08-731-045-2	Sequence 18, Appli
163	23.8	5	412	4	US-09-026-408-3	Sequence 5, Appli
164	23.8	5	415	3	US-08-208-007A-12	Sequence 5, Appli
165	23.8	5	417	4	US-08-974-691-4	Sequence 1, Appli
166	23.8	5	417	3	US-09-100-193-2	
167	23.8	5	436	4	US-08-348-817A-4	
168	23.8	5	436	4	US-09-026-408-8	
169	23.8	5	437	1	US-08-705-771-18	
170	23.8	5			US-08-716-873-5	
171	23.8	5			US-09-368-431-5	
172	23.8	5			US-08-764-343-1	
173	23.8	5				
174	23.8	5	437	2	US-08-989-925-3	Sequence 3, Appli
175	23.8	5	439	4	US-08-716-873-2	Sequence 2, Appli
176	23.8	5	439	4	US-08-716-873-4	Sequence 4, Appli
177	23.8	5	439	4	US-09-368-431-2	Sequence 2, Appli
178	23.8	5	439	4	US-09-368-431-4	Sequence 4, Appli
179	23.8	5	444	1	US-09-252-292C-27	Sequence 27, Appl
180	23.8	5	445	1	US-08-353-400-33	Sequence 33, Appl
181	23.8	5	447	6	5455030-1	Patent No. 5455030
182	23.8	5	450	4	US-09-268-992-41	Sequence 41, Appl
183	23.8	5	464	1	US-08-353-400-36	Sequence 36, Appl
184	23.8	5	464	2	US-08-553-619B-9	Sequence 9, Appli
185	23.8	5	466	4	US-09-268-992-39	Sequence 39, Appl
186	23.8	5	467	4	US-09-086-483A-6	Sequence 6, Appli
187	23.8	5	467	2	US-08-495-484-12	Sequence 12, Appl
188	23.8	5	468	2	US-08-303-569B-7	Sequence 7, Appli
189	23.8	5	468	2	US-08-116-247-7	Sequence 7, Appli
190	23.8	5	468	2	US-09-013-895A-2	Sequence 2, Appli
191	23.8	5	469	4	US-08-968-751-2	Sequence 2, Appli
192	23.8	5	469	4	US-09-052-089A-1	Sequence 1, Appli
193	23.8	5	473	3	US-09-171-945-131	Sequence 131, App
194	23.8	5	477	4	US-09-446-504-19	Sequence 19, Appl
195	23.8	5	479	4	US-09-712-266-19	Sequence 19, Appl
196	23.8	5	479	1	US-08-484-105-10	Sequence 10, Appl
197	23.8	5	479	1	US-08-484-106-10	Sequence 10, Appl
198	23.8	5	480	2	US-08-828-488-8	Sequence 8, Appli
199	23.8	5	508	1	US-08-724-281-1	Sequence 1, Appli
200	23.8	5	516	2	US-09-201-641-6	Sequence 6, Appli
201	23.8	5	524	1	US-08-624-125-2	Sequence 2, Appli
202	23.8	5	524	1	US-08-624-125-21	Sequence 21, Appl
203	23.8	5	533	3	US-08-770-544-4	Sequence 4, Appli
204	23.8	5	535	3	US-08-813-574-2	Sequence 2, Appli
205	23.8	5	538	1	US-08-541-759B-2	Sequence 2, Appli
206	23.8	5	580	1	US-08-309-512-6	Sequence 6, Appli
207	23.8	5	580	5	PCT-US92-08756A-6	Sequence 6, Appli
208	23.8	5	585	1	US-08-477-674-10	Sequence 10, Appl
209	23.8	5	585	1	US-08-473-791-10	Sequence 10, Appl
210	23.8	5	585	3	US-08-316-714-10	Sequence 10, Appl
211	23.8	5	585	1	US-08-473-673-10	Sequence 10, Appl
212	23.8	5	599	3	US-08-442-542-18	Sequence 18, Appl
213	23.8	5	599	3	US-08-765-469-18	Sequence 18, Appl
214	23.8	5	626	2	US-08-596-300A-7	Sequence 7, Appli
215	23.8	5	626	2	US-08-596-300A-14	Sequence 14, Appl
216	23.8	5	630	4	US-08-609-236-2	Sequence 2, Appli
217	23.8	5	638	1	US-09-704-449-2	Sequence 2, Appli
218	23.8	5	639	1	US-09-228-986-74	Sequence 74, Appl
219	23.8	5	639	4	US-09-466-390-2	Sequence 2, Appli
220	23.8	5	639	1	US-08-470-950-2	Sequence 2, Appli
221	23.8	5	639	1	US-08-467-781-2	Sequence 2, Appli
222	23.8	5	639	3	US-08-195-487-2	Sequence 2, Appli
223	23.8	5	639	2	US-08-483-924-2	Sequence 2, Appli
224	23.8	5	639	5	PCT-US93-06160-2	Sequence 25, Appl
225	23.8	5	643	4	US-09-178-252-25	Sequence 60, Appl
226	23.8	5	647	4	US-09-423-439-60	Sequence 2, Appli
227	23.8	5	670	4	US-08-980-080-2	Sequence 13, Appl
228	23.8	5	685	3	US-08-878-989-1	Sequence 16, Appl
229	23.8	5	685	4	US-08-836-567-8	Sequence 8, Appli
230	23.8	5	685	4	US-07-731-157A-7	Sequence 7, Appli
231	23.8	5	713	4	US-08-541-780-7	Sequence 4, Appli
232	23.8	5	726	3	US-08-909-994-4	Sequence 23, Appl
233	23.8	5	726	4	US-08-380-182-23	Sequence 2, Appli
234	23.8	5	756	4	US-09-291-839-2	Sequence 9, Appli
235	23.8	5	767	4	US-08-395-580-2	Sequence 3, Appli
236	23.8	5	774	1	PCT-US95-02792-2	Sequence 2, Appli
237	23.8	5	774	3	US-08-911-853-17	Sequence 2, Appli
238	23.8	5	799	2	US-09-479-409-17	Sequence 17, Appl
239	23.8	5	820	1	US-09-479-453-17	Sequence 17, Appl

Left table:

No.				Patent/Ref Number		Description
247	948	23.8	5	US-08-698-551-14	1	Sequence 14, Appl
248	948	23.8	5	US-08-602-228-14	2	Sequence 14, Appl
249	948	23.8	5	US-08-533-901B-14	2	Sequence 14, Appl
250	948	23.8	5	US-08-839-032A-14	2	Sequence 14, Appl
251	948	23.8	5	US-08-839-031A-14	2	Sequence 14, Appl
252	948	23.8	5	PCT-US95-12724-14	5	Sequence 14, Appl
253	956	23.8	5	US-08-897-443-3	1	Sequence 3, Appli
254	966	23.8	5	US-08-571-758-2	1	Sequence 2, Appli
255	966	23.8	5	US-08-909-984A-2	1	Sequence 2, Appli
256	966	23.8	5	US-08-909-983-2	1	Sequence 2, Appli
257	1050	23.8	5	US-09-045-632-49	3	Sequence 49, Appl
258	1050	23.8	5	US-09-045-632-50	3	Sequence 50, Appl
259	1066	23.8	5	US-09-004-838-24	4	Sequence 24, Appl
260	1141	23.8	5	US-08-131-365B-54	1	Sequence 54, Appl
261	1141	23.8	5	US-08-668-123-54	4	Sequence 54, Appl
262	1186	23.8	5	US-09-178-252-23	4	Sequence 23, Appl
263	1203	23.8	5	US-09-075-272-4	4	Sequence 4, Appli
264	1222	23.8	5	US-09-004-838-119	5	Sequence 119, App
265	1227	23.8	5	US-08-448-170-8	2	Sequence 8, Appli
266	1227	23.8	5	US-08-961-803-9	1	Sequence 9, Appli
267	1275	23.8	5	US-09-120-513-2	3	Sequence 2, Appli
268	1275	23.8	5	US-09-450-105-2	4	Sequence 2, Appli
269	1279	23.8	5	US-08-784-649A-2	2	Sequence 2, Appli
270	1280	23.8	5	US-08-583-276-19	2	Sequence 19, Appl
271	1280	23.8	5	US-08-752-447-2	2	Sequence 2, Appli
272	1280	23.8	5	5206352-4	6	Patent No. 5206352
273	1323	23.8	5	US-09-004-838-90	2	Sequence 90, Appl
274	1434	23.8	5	US-08-540-406-10	3	Sequence 10, Appl
275	1434	23.8	5	US-08-656-055-10	4	Sequence 10, Appl
276	1434	23.8	5	US-08-954-668-10	4	Sequence 10, Appl
277	1434	23.8	5	PCT-US95-13233-10	5	Sequence 10, Appl
278	1581	23.8	5	US-09-110-517-2	2	Sequence 2, Appli
279	1786	23.8	5	US-08-477-451-16	2	Sequence 16, Appl
280	1786	23.8	5	US-08-973-462-8	4	Sequence 8, Appli
281	1817	23.8	5	US-09-004-838-125	4	Sequence 125, App
282	1872	23.8	5	5386025-6	6	Patent No. 5386025
283	1873	23.8	5	US-08-435-675B-4	1	Sequence 4, Appli
284	1873	23.8	5	US-08-336-257A-7	1	Sequence 7, Appli
285	1940	23.8	5	US-08-644-271-30	2	Sequence 30, Appl
286	2441	23.8	5	US-08-194-468-2	1	Sequence 2, Appli
287	2441	23.8	5	US-08-961-739-2	3	Sequence 2, Appli
288	2516	23.8	5	US-08-374-077C-2	4	Sequence 2, Appli
289	2802	23.8	5	US-08-895-590-2	2	Sequence 2, Appli
290	3072	23.8	5	US-09-542-331-1	4	Sequence 1, Appli
291	3079	23.8	5	US-09-413-814-93	4	Sequence 93, Appl
292	3118	23.8	5	US-09-413-814-80	4	Sequence 80, Appl
293	3118	23.8	5	US-08-457-273B-8	1	Sequence 8, Appli
294	3119	23.8	5	US-08-246-982A-16	1	Sequence 16, Appl
295	3119	23.8	5	US-08-453-265-16	1	Sequence 16, Appl
296	3144	23.8	5	US-08-246-982A-6	1	Sequence 6, Appli
297	3144	23.8	5	US-08-453-265-6	1	Sequence 6, Appli
298	3144	23.8	5	US-08-457-273B-42	2	Sequence 42, Appl
299	3144	23.8	5	US-09-413-814-21	4	Sequence 21, Appl
300	3144	23.8	5	US-08-556-419-21	4	Sequence 15, Appl
301	3144	23.8	5	US-09-041-886-15	4	Sequence 15, Appl
302	7257	23.8	5	US-09-335-409-5	3	Sequence 5, Appli
303	7257	23.8	5	US-09-568-102-5	4	Sequence 5, Appli
304	7257	23.8	5	US-09-567-969-5	4	Sequence 5, Appli
305	7257	23.8	5	US-09-568-480-5	4	Sequence 5, Appli
306	7257	23.8	5	US-09-568-486-5	4	Sequence 5, Appli
307	7257	23.8	5	US-09-568-472-5	4	Sequence 4, Appli
308		19.0	4	US-08-477-727A-54	1	Sequence 54, Appl
309	6	19.0	4	US-08-724-548-50	2	Sequence 50, Appl
310	6	19.0	4	US-08-724-548-51	2	Sequence 51, Appl
311	6	19.0	4	US-07-978-674B-50	3	Sequence 50, Appl
312	6	19.0	4	US-07-978-674B-51	3	Sequence 51, Appl
313	6	19.0	4	US-09-156-580-6	1	Sequence 6, Appli
314	6	19.0	4	US-09-156-579C-8	6	Sequence 8, Appli
315	6	19.0	4	5204326-147	6	Patent No. 5204326
316	7	19.0	4	US-08-470-837-8	1	Sequence 8, Appli
317	7	19.0	4	US-08-632-598-25	2	Sequence 25, Appl
318	7	19.0	4	US-09-173-941-16	2	Sequence 16, Appl
319	7	19.0	4	US-09-173-941-100	2	Sequence 100, App
	7	19.0	4	US-09-231-240-25	4	Sequence 25, Appl

Right table:

No.				Patent/Ref Number			Description
320	7	4	19.0	4	US-08-868-452-8		Sequence 8, Appli
321	7	6	19.0	4	5204326-110		Patent No. 5204326
322	8	2	19.0	4	US-08-646-360-141		Sequence 141, App
323	8	2	19.0	4	US-08-792-553-13		Sequence 13, Appl
324	8	5	19.0	4	US-09-136-389-141		Sequence 141, App
325	8	6	19.0	4	5449756-6		Patent No. 5449756
326	9	1	19.0	4	US-08-619-645-1		Sequence 1, Appli
327	9	1	19.0	4	US-08-634-493-1		Sequence 1, Appli
328	9	3	19.0	4	US-09-079-447-1		Sequence 1, Appli
329	9	4	19.0	4	US-08-660-092-72		Sequence 72, Appl
330	10	1	19.0	4	US-07-864-475A-11		Sequence 11, Appl
331	10	1	19.0	4	US-08-218-026-38		Sequence 38, Appl
332	10	2	19.0	4	US-08-724-548-9		Sequence 9, Appli
333	10	2	19.0	4	US-08-724-548-10		Sequence 10, Appl
334	10	2	19.0	4	US-08-724-548-11		Sequence 11, Appl
335	10	2	19.0	4	US-08-724-548-12		Sequence 12, Appl
336	10	2	19.0	4	US-08-724-548-13		Sequence 13, Appl
337	10	2	19.0	4	US-08-724-548-14		Sequence 14, Appl
338	10	2	19.0	4	US-08-724-548-15		Sequence 15, Appl
339	10	2	19.0	4	US-08-653-632-38		Sequence 38, Appl
340	10	2	19.0	4	US-08-637-759B-192		Sequence 192, App
341	10	2	19.0	4	US-08-468-249A-11		Sequence 11, Appl
342	10	3	19.0	4	US-08-596-257A-8		Sequence 8, Appli
343	10	3	19.0	4	US-08-871-355A-192		Sequence 192, App
344	10	3	19.0	4	US-07-978-674B-9		Sequence 9, Appli
345	10	3	19.0	4	US-07-978-674B-10		Sequence 10, Appl
346	10	3	19.0	4	US-07-978-674B-11		Sequence 11, Appl
347	10	3	19.0	4	US-07-978-674B-12		Sequence 12, Appl
348	10	3	19.0	4	US-07-978-674B-13		Sequence 13, Appl
349	10	3	19.0	4	US-07-978-674B-14		Sequence 14, Appl
350	10	3	19.0	4	US-07-978-674B-15		Sequence 15, Appl
351	10	4	19.0	4	US-08-860-339-8		Sequence 8, Appli
352	10	4	19.0	4	US-08-939-853A-16		Sequence 16, Appl
353	10	4	19.0	4	US-09-201-945-192		Sequence 192, App
354	10	6	19.0	4	5204326-111		Patent No. 5204326
355	12	1	19.0	4	US-08-309-512-56		Sequence 56, Appl
356	12	2	19.0	4	US-08-479-614-10		Sequence 10, Appl
357	13	1	19.0	4	US-08-209-525-38		Sequence 38, Appl
358	13	1	19.0	4	US-08-619-645-5		Sequence 5, Appli
359	13	1	19.0	4	US-08-619-645-6		Sequence 6, Appli
360	13	1	19.0	4	US-08-634-493-5		Sequence 5, Appli
361	13	3	19.0	4	US-08-634-493-6		Sequence 6, Appli
362	13	4	19.0	4	US-08-423-646A-54		Sequence 54, Appl
363	13	5	19.0	4	PCT-US93-05647-11		Sequence 11, Appl
364	15	1	19.0	4	US-08-268-251-27		Sequence 27, Appl
365	15	5	19.0	4	PCT-US93-01112-27		Sequence 27, Appl
366	16	1	19.0	4	US-07-940-861-7		Sequence 7, Appli
367	16	2	19.0	4	US-08-459-512-7		Sequence 7, Appli
368	16	2	19.0	4	US-08-459-657-7		Sequence 7, Appli
369	16	2	19.0	4	US-08-460-132-7		Sequence 7, Appli
370	16	3	19.0	4	US-09-133-774-5		Sequence 5, Appli
371	16	4	19.0	4	US-08-312-202B-4		Sequence 4, Appli
372	16	5	19.0	4	US-09-303-862-5		Sequence 5, Appli
373	16	3	19.0	4	US-09-079-347-4		Sequence 7, Appli
374	16	3	19.0	4	US-09-075-725-4		Sequence 7, Appli
375	16	4	19.0	4	US-09-171-705-40		Sequence 5, Appli
376	16	4	19.0	4	US-09-171-705-41		Sequence 4, Appli
377	16	4	19.0	4	US-08-809-646-4		Sequence 40, Appl
378	16	5	19.0	4	PCT-US92-02050-7		Sequence 41, Appli
379	16	5	19.0	4	PCT-US95-12433-4		Sequence 4, Appli
380	17	1	19.0	4	US-08-850-910A-15		Sequence 7, Appli
381	17	6	19.0	4	5185441-3		Patent No. 5185441
382	18	1	19.0	4	US-08-476-405A-1		Sequence 15, Appl
383	18	1	19.0	4	US-08-499-523-15		Sequence 1, Appli
384	18	4	19.0	4	US-09-128-345-15		Sequence 15, Appl
385	19	1	19.0	4	US-08-111-365B-44		Sequence 15, Appl
386	19	2	19.0	4	US-07-678-123-44		Sequence 44, Appl
387	19	1	19.0	4	US-07-678-974D-19		Sequence 44, Appl
388	20	1	19.0	4	US-08-476-405A-2		Sequence 19, Appl
389	20	1	19.0	4	US-08-850-910A-34		Sequence 2, Appli
390	20	2	19.0	4	US-08-945-168-24		Sequence 34, Appl
391	20	2	19.0	4	US-08-737-226-11		Sequence 24, Appl
392	20	4	19.0	4	US-09-324-910-2		Sequence 11, Appl
							Sequence 2, Appli

393	4	19.0	21	1	US-07-851-976B-1	Sequence 1, Appli
394	4	19.0	21	1	US-07-908-455A-7	Sequence 7, Appli
395	4	19.0	21	1	US-08-134-120-33	Sequence 33, Appl
396	4	19.0	21	1	US-08-291-609-1	Sequence 1, Appli
397	4	19.0	21	1	US-08-465-325-33	Sequence 33, Appl
398	4	19.0	21	1	US-08-401-136-1	Sequence 1, Appli
399	4	19.0	21	3	US-08-850-554-1	Sequence 1, Appl
400	4	19.0	21	4	US-09-115-737-33	Sequence 33, Appl
401	4	19.0	21	6	5182195-5	Patent No. 5182195
402	4	19.0	22	3	US-08-833-553-6	Sequence 6, Appli
403	4	19.0	22	4	US-09-418-222-6	Sequence 6, Appli
404	4	19.0	22	4	US-08-706-344C-8	Sequence 8, Appli
405	4	19.0	22	1	US-09-439-313-331	Sequence 331, App
406	4	19.0	23	1	US-07-656-566-1	Sequence 1, Appl
407	4	19.0	23	1	US-08-268-251-57	Sequence 57, Appl
408	4	19.0	23	2	US-08-146-028-28	Sequence 28, Appl
409	4	19.0	23	2	US-08-146-028-167	Sequence 167, App
410	4	19.0	23	4	US-08-723-425A-28	Sequence 28, Appl
411	4	19.0	23	4	US-08-723-425A-167	Sequence 167, App
412	4	19.0	23	4	US-09-112-206-28	Sequence 28, Appl
413	4	19.0	23	4	US-09-112-206-167	Sequence 167, App
414	4	19.0	23	5	PCT-US93-01112-57	Sequence 57, Appl
415	4	19.0	24	3	US-08-701-382-8	Sequence 8, Appli
416	4	19.0	24	3	US-08-894-699-35	Sequence 35, Appl
417	4	19.0	24	4	US-09-444-410-35	Sequence 35, Appl
418	4	19.0	24	4	US-08-788-820-8	Sequence 8, Appli
419	4	19.0	25	2	US-08-632-598-43	Sequence 43, Appl
420	4	19.0	25	2	US-08-632-598-44	Sequence 44, Appl
421	4	19.0	25	2	US-08-632-598-45	Sequence 45, Appl
422	4	19.0	25	2	US-08-632-598-46	Sequence 46, Appl
423	4	19.0	25	2	US-08-765-452-15	Sequence 15, Appl
424	4	19.0	25	4	US-09-231-240-43	Sequence 43, Appl
425	4	19.0	25	4	US-09-231-240-44	Sequence 44, Appl
426	4	19.0	25	4	US-09-231-240-45	Sequence 45, Appl
427	4	19.0	25	4	US-09-231-240-46	Sequence 46, Appl
428	4	19.0	26	1	US-07-778-847-2	Sequence 2, Appli
429	4	19.0	26	2	US-08-856-006-2	Sequence 2, Appli
430	4	19.0	26	2	US-08-620-151-3	Sequence 3, Appli
431	4	19.0	26	2	US-08-850-910A-3	Sequence 3, Appli
432	4	19.0	26	2	US-09-122-642-2	Sequence 2, Appli
433	4	19.0	26	3	US-08-851-843A-145	Sequence 145, App
434	4	19.0	26	4	US-08-974-549A-263	Sequence 263, App
435	4	19.0	26	4	US-08-854-050-145	Sequence 145, App
436	4	19.0	26	4	US-09-524-014-2	Sequence 2, Appli
437	4	19.0	26	4	US-09-430-323-145	Sequence 145, App
438	4	19.0	27	2	US-08-446-692-7	Sequence 7, Appli
439	4	19.0	27	2	US-08-488-351A-7	Sequence 7, Appli
440	4	19.0	27	3	US-08-749-816-6	Sequence 6, Appli
441	4	19.0	27	3	US-08-318-794-29	Sequence 29, Appl
442	4	19.0	27	3	US-09-100-409A-43	Sequence 43, Appl
443	4	19.0	27	3	US-08-851-843A-150	Sequence 150, App
444	4	19.0	27	4	US-08-974-549A-270	Sequence 270, App
445	4	19.0	27	4	US-08-854-050-150	Sequence 150, App
446	4	19.0	27	4	US-09-144-914-10	Sequence 10, Appl
447	4	19.0	27	4	US-09-130-323-150	Sequence 150, App
448	4	19.0	27	4	US-08-470-106-29	Sequence 29, Appl
449	4	19.0	29	5	PCT-US95-13841-10	Sequence 10, Appl
450	4	19.0	29	1	US-07-754-958-6	Sequence 6, Appli
451	4	19.0	29	1	US-07-754-947-6	Sequence 6, Appli
452	4	19.0	29	1	US-09-227-357-402	Sequence 402, App
453	4	19.0	30	1	US-08-305-799A-10	Sequence 10, Appl
454	4	19.0	30	2	US-08-753-629A-13	Sequence 13, Appl
455	4	19.0	30	2	US-08-945-168-107	Sequence 107, App
456	4	19.0	30	4	US-09-230-405-11	Sequence 11, Appl
457	4	19.0	31	1	US-08-190-802A-236	Sequence 236, App
458	4	19.0	31	1	US-08-732-751-3	Sequence 3, Appli
459	4	19.0	31	1	US-08-732-751-4	Sequence 49, Appli
460	4	19.0	31	3	US-07-401-432-49	Sequence 49, Appl
461	4	19.0	31	3	US-08-477-346-236	Sequence 236, App
462	4	19.0	31	4	US-08-473-089-236	Sequence 236, App
463	4	19.0	31	6	5204326-138	Patent No. 5204326
464	4	19.0	31	6	5204326-139	Patent No. 5204326
465	4	19.0	31	6	5204326-141	Patent No. 5204326
466	4	19.0	31	6	5204326-142	Patent No. 5204326
467	4	19.0	31	6	5204326-144	Patent No. 5204326
468	4	19.0	32	1	US-07-754-958-5	Sequence 5, Appli
469	4	19.0	32	1	US-07-755-947-5	Sequence 5, Appli
470	4	19.0	32	1	US-07-952-735A-3	Sequence 3, Appli
471	4	19.0	32	1	US-07-952-735A-4	Sequence 4, Appli
472	4	19.0	32	1	US-08-305-799A-1	Sequence 1, Appli
473	4	19.0	32	2	US-08-526-583-5	Sequence 5, Appli
474	4	19.0	32	2	US-08-526-583-6	Sequence 6, Appli
475	4	19.0	32	2	US-08-526-583-7	Sequence 7, Appli
476	4	19.0	32	2	US-08-526-583-8	Sequence 8, Appli
477	4	19.0	32	2	US-08-595-868C-41	Sequence 41, Appl
478	4	19.0	32	3	US-08-954-915A-44	Sequence 44, Appl
479	4	19.0	32	3	US-08-954-915A-45	Sequence 45, Appl
480	4	19.0	32	3	US-08-484-322-67	Sequence 67, Appl
481	4	19.0	32	3	US-07-401-432-10	Sequence 10, Appl
482	4	19.0	32	3	US-07-401-432-50	Sequence 50, Appl
483	4	19.0	32	3	US-07-401-432-51	Sequence 51, Appl
484	4	19.0	32	4	US-08-905-223-415	Sequence 415, App
485	4	19.0	32	4	US-09-139-819A-41	Sequence 41, Appl
486	4	19.0	32	6	5183802-5	Patent No. 5183302
487	4	19.0	32	6	5204326-143	Patent No. 5204326
488	4	19.0	32	6	5208144-17	Patent No. 5208144
489	4	19.0	33	2	US-07-776-272-9	Sequence 9, Appli
490	4	19.0	33	2	US-08-078-311-26	Sequence 26, Appl
491	4	19.0	33	2	US-08-460-402-26	Sequence 26, Appl
492	4	19.0	33	3	US-09-066-046-9	Sequence 9, Appli
493	4	19.0	34	1	US-07-952-735A-1	Sequence 1, Appli
494	4	19.0	34	1	US-07-952-735A-2	Sequence 2, Appli
495	4	19.0	34	4	US-08-974-549A-38	Sequence 38, Appl
496	4	19.0	34	4	US-08-482-918-79	Sequence 79, Appl
497	4	19.0	34	4	US-09-224-681-79	Sequence 79, Appl
498	4	19.0	34	4	US-08-336-728A-79	Sequence 79, Appl
499	4	19.0	34	6	5204326-2	Patent No. 5204326
500	4	19.0	35	2	US-09-227-357-183	Sequence 183, App
501	4	19.0	36	1	US-08-487-890A-32	Sequence 32, Appl
502	4	19.0	36	2	US-08-478-435-32	Sequence 32, Appl
503	4	19.0	36	2	US-08-337-483-32	Sequence 32, Appl
504	4	19.0	36	2	US-08-478-373-32	Sequence 32, Appl
505	4	19.0	36	3	US-08-474-671-32	Sequence 32, Appl
506	4	19.0	36	3	US-08-483-577A-32	Sequence 32, Appl
507	4	19.0	36	4	US-08-897-438-32	Sequence 32, Appl
508	4	19.0	37	2	US-08-637-654-32	Sequence 32, Appl
509	4	19.0	37	2	US-08-446-692-15	Sequence 15, Appl
510	4	19.0	37	2	US-08-488-351A-15	Sequence 15, Appl
511	4	19.0	37	4	US-08-482-918-80	Sequence 80, Appl
512	4	19.0	37	4	US-09-224-681-80	Sequence 80, Appl
513	4	19.0	38	1	US-08-336-728A-80	Sequence 80, Appl
514	4	19.0	38	1	US-08-176-500-46	Sequence 46, Appl
515	4	19.0	38	1	US-08-471-052A-46	Sequence 46, Appl
516	4	19.0	38	2	US-08-189-331-46	Sequence 46, Appl
517	4	19.0	38	2	US-08-471-939-46	Sequence 46, Appl
518	4	19.0	38	2	US-08-471-800-46	Sequence 46, Appl
519	4	19.0	39	1	US-08-471-068-46	Sequence 46, Appl
520	4	19.0	39	1	US-08-324-301-6	Sequence 6, Appli
521	4	19.0	39	3	US-08-258-287B-46	Sequence 46, Appl
522	4	19.0	39	3	US-08-368-704C-45	Sequence 45, Appl
523	4	19.0	40	1	US-08-188-228-12	Sequence 12, Appl
524	4	19.0	40	1	US-08-188-228-18	Sequence 18, Appl
525	4	19.0	40	1	US-08-188-228-24	Sequence 24, Appl
526	4	19.0	40	1	US-08-332-643-12	Sequence 12, Appl
527	4	19.0	40	1	US-08-332-643-18	Sequence 18, Appl
528	4	19.0	40	1	US-08-332-643-24	Sequence 24, Appl
529	4	19.0	40	1	US-08-332-638-12	Sequence 12, Appl
530	4	19.0	40	1	US-08-332-638-18	Sequence 18, Appl
531	4	19.0	40	2	US-08-332-638-24	Sequence 24, Appl
532	4	19.0	40	2	US-08-031-538-52	Sequence 52, Appl
533	4	19.0	40	4	US-08-968-685A-12	Sequence 12, Appl
534	4	19.0	40	4	US-08-988-856B-19	Sequence 19, Appl
535	4	19.0	41	3	US-08-630-915A-83	Sequence 83, Appl
536	4	19.0	41	4	US-08-984-277-8	Sequence 8, Appli
537	4	19.0	41	4	US-09-156-580-11	Sequence 11, Appl
538	4	19.0	41	4	US-08-988-856B-18	Sequence 18, Appl

Left column

#					Accession	Description
539	4	19.0	41	4	US-09-156-579C-13	Sequence 13, Appl
540	4	19.0	42	2	US-08-476-866-23	Sequence 23, Appl
541	4	19.0	42	4	US-09-156-580-10	Sequence 10, Appl
542	4	19.0	42	4	US-09-156-579C-12	Sequence 6, Appl
543	4	19.0	43	1	US-08-325-547-6	Sequence 6, Appli
544	4	19.0	43	4	US-08-968-685A-1	Sequence 9, Appli
545	4	19.0	43	4	US-09-156-580-9	Sequence 12, Appl
546	4	19.0	43	4	US-09-156-580-12	Sequence 13, Appl
547	4	19.0	43	4	US-09-156-580-13	Sequence 11, Appl
548	4	19.0	43	4	US-09-156-579C-11	Sequence 14, Appl
549	4	19.0	43	4	US-09-156-579C-14	Sequence 21, Appl
550	4	19.0	43	4	US-09-156-579C-15	Sequence 20, Appl
551	4	19.0	44	5	PCT-US92-00282-21	Sequence 11, Appli
552	4	19.0	45	2	US-07-689-693B-20	Sequence 6, Appli
553	4	19.0	45	4	US-08-935-450-11	Sequence 10, Appl
554	4	19.0	46	1	US-09-257-218-6	Sequence 15, Appl
555	4	19.0	47	4	US-09-311-760-6	Sequence 12, Appl
556	4	19.0	47	5	US-08-222-616-10	Sequence 10, Appl
557	4	19.0	47	4	US-08-939-853A-15	Sequence 7, Appli
558	4	19.0	48	3	US-08-963-851-12	Sequence 59, Appl
559	4	19.0	48	4	PCT-US95-04228-10	Sequence 16, Appl
560	4	19.0	48	4	US-09-222-373-5	Sequence 18, Appl
561	4	19.0	48	4	US-09-222-373-7	Sequence 19, Appl
562	4	19.0	48	4	US-08-996-679-59	Sequence 5, Appli
563	4	19.0	48	4	US-08-945-997C-16	Sequence 7, Appli
564	4	19.0	48	4	US-08-945-997C-18	Sequence 69, Appli
565	4	19.0	48	4	US-08-945-997C-19	Sequence 5, Appli
566	4	19.0	48	4	US-09-001-511-5	Sequence 78, Appl
567	4	19.0	48	4	US-09-001-511-7	Sequence 62, Appl
568	4	19.0	48	3	US-09-113-977C-69	Sequence 8, Appli
569	4	19.0	48	4	US-09-510-616-5	Sequence 8, Appli
570	4	19.0	48	3	US-09-510-616-7	Sequence 10, Appl
571	4	19.0	48	4	US-09-248-074-78	Sequence 46, Appl
572	4	19.0	48	4	US-09-187-859-62	Sequence 46, Appl
573	4	19.0	50	3	US-09-222-373-8	Sequence 289, App
574	4	19.0	50	4	US-09-001-511-8	Sequence 79, Appl
575	4	19.0	51	4	US-09-510-616-8	Sequence 326, App
576	4	19.0	51	3	US-09-174-465D-10	Sequence 52, Appl
577	4	19.0	51	4	US-09-001-511-46	Sequence 52, Appl
578	4	19.0	51	4	US-09-510-616-46	Sequence 73, Appl
579	4	19.0	52	4	US-08-858-207A-289	Sequence 22, Appl
580	4	19.0	53	3	US-08-630-916A-79	Sequence 73, Appl
581	4	19.0	53	3	US-08-905-223-326	Sequence 73, Appl
582	4	19.0	55	4	US-08-706-741B-52	Sequence 73, Appl
583	4	19.0	55	2	US-08-924-695A-52	Sequence 73, Appl
584	4	19.0	56	1	US-08-485-455D-73	Sequence 73, Appl
585	4	19.0	56	4	US-08-592-406-22	Sequence 73, Appl
586	4	19.0	56	2	US-08-482-130C-73	Sequence 73, Appl
587	4	19.0	56	3	US-08-484-211C-73	Sequence 73, Appl
588	4	19.0	56	3	US-08-906-769-73	Sequence 186, App
589	4	19.0	56	3	US-08-906-616-73	Sequence 187, App
590	4	19.0	56	3	US-08-817-795-73	Sequence 188, App
591	4	19.0	56	4	US-08-639-075A-73	Sequence 189, Appl
592	4	19.0	56	4	US-08-485-443B-73	Sequence 73, Appl
593	4	19.0	56	4	US-09-012-431-73	Sequence 58, Appl
594	4	19.0	56	4	US-09-012-692-73	Sequence 7, Appli
595	4	19.0	56	4	US-08-711-417C-186	Sequence 7, Appli
596	4	19.0	56	4	US-08-711-417C-187	Sequence 7, Appli
597	4	19.0	56	4	US-08-711-417C-188	Sequence 3, Appli
598	4	19.0	56	5	US-08-711-417C-189	Sequence 3, Appli
599	4	19.0	56	5	US-08-906-613-73	Sequence 12, Appl
600	4	19.0	56	4	US-09-187-789-58	Sequence 115, App
601	4	19.0	56	5	PCT-US95-14442A-73	Sequence 115, App
602	4	19.0	56	3	US-08-955-937A-7	Sequence 58, Appl
603	4	19.0	57	4	US-09-300-985-7	Sequence 7, Appli
604	4	19.0	59	1	US-08-488-961-3	Sequence 3, Appli
605	4	19.0	59	4	US-08-973-297-3	Sequence 3, Appli
606	4	19.0	59	5	PCT-US96-06511-3	Sequence 3, Appli
607	4	19.0	61	2	US-08-687-702-12	Sequence 12, Appl
608	4	19.0	62	1	US-08-464-531-115	Sequence 115, App
609	4	19.0	62	4	US-08-461-598-115	Sequence 115, App
610	4	19.0	62	3	US-08-122-137-115	Sequence 115, App

Right column

#					Accession	Description
612	4	19.0	64	2	US-08-765-179B-23	Sequence 23, App
613	4	19.0	64	3	US-08-480-640A-194	Sequence 194, App
614	4	19.0	64	4	US-08-686-968C-194	Sequence 194, App
615	4	19.0	64	4	US-08-488-237A-194	Sequence 194, App
616	4	19.0	64	4	US-08-375-992A-194	Sequence 194, App
617	4	19.0	65	4	US-09-197-636-6	Sequence 6, Appli
618	4	19.0	65	4	US-09-461-697-447	Sequence 447, App
619	4	19.0	65	4	US-09-235-451-9	Sequence 9, Appli
620	4	19.0	66	1	US-07-626-618A-16	Sequence 16, Appl
621	4	19.0	66	1	US-07-928-611-16	Sequence 16, Appl
622	4	19.0	66	1	US-07-333-977-16	Sequence 16, Appl
623	4	19.0	66	2	US-08-487-811A-16	Sequence 16, Appl
624	4	19.0	66	5	US-09-060-694-16	Sequence 16, Appl
625	4	19.0	67	4	PCT-US93-07370-16	Sequence 379, App
626	4	19.0	67	4	US-08-905-223-379	Sequence 435, App
627	4	19.0	67	4	US-08-905-223-435	Sequence 445, App
628	4	19.0	67	4	US-09-461-697-445	Sequence 4, Appli
629	4	19.0	68	3	US-09-243-675-4	Sequence 9, Appli
630	4	19.0	68	4	US-09-418-222-9	Sequence 15, Appl
631	4	19.0	68	1	US-08-833-553-9	Sequence 115, App
632	4	19.0	69	2	US-08-963-851-15	Sequence 115, App
633	4	19.0	69	3	US-08-056-200-115	Sequence 13, Appl
634	4	19.0	69	1	US-08-800-644-115	Sequence 13, Appl
635	4	19.0	69	1	US-08-851-843A-13	Sequence 15, Appl
636	4	19.0	69	2	US-08-854-050-13	Sequence 54, Appl
637	4	19.0	69	1	US-09-430-323-13	Sequence 54, Appl
638	4	19.0	70	1	US-08-691-641-15	Sequence 12, Appl
639	4	19.0	70	2	US-08-724-548-54	Sequence 1, Appl
640	4	19.0	70	1	US-07-978-674B-54	Sequence 6, Appli
641	4	19.0	73	1	US-07-689-693B-3	Sequence 8, Appli
642	4	19.0	73	1	US-08-624-123-12	Sequence 10, Appl
643	4	19.0	73	5	US-08-518-967-1	Sequence 25, Appl
644	4	19.0	76	5	PCT-US96-05262-13	Sequence 7, Appli
645	4	19.0	76	1	PCT-US95-13658-6	Sequence 59, Appl
646	4	19.0	77	2	US-08-014-153D-40	Sequence 59, Appl
647	4	19.0	77	3	US-08-570-227A-8	Sequence 7, Appli
648	4	19.0	79	3	US-09-077-991-10	Sequence 59, Appl
649	4	19.0	79	1	US-09-143-776-6	Sequence 7, Appli
650	4	19.0	82	4	US-08-476-405A-25	Sequence 126, App
651	4	19.0	82	4	US-09-382-155-7	Sequence 8, Appli
652	4	19.0	82	4	US-08-482-918-59	Sequence 88, Appl
653	4	19.0	82	4	US-09-224-681-59	Sequence 27, Appl
654	4	19.0	82	7	US-09-074-044A-7	Sequence 14, Appl
655	4	19.0	85	1	US-08-336-728A-59	Sequence 27, Appl
656	4	19.0	85	3	US-08-236-886-7	Sequence 2, Appli
657	4	19.0	86	4	US-09-203-623-4	Sequence 13, Appl
658	4	19.0	86	1	US-08-053-451B-126	Sequence 3, Appli
659	4	19.0	86	4	US-09-382-155-8	Sequence 433, App
660	4	19.0	87	4	US-08-074-044A-8	Sequence 8, Appli
661	4	19.0	88	4	US-08-332-562A-88	Sequence 88, Appl
662	4	19.0	88	4	US-08-562-114B-27	Sequence 27, Appl
663	4	19.0	88	3	US-09-053-197A-14	Sequence 14, Appl
664	4	19.0	88	3	US-08-729-594A-27	Sequence 27, Appl
665	4	19.0	88	4	US-09-085-761A-14	Sequence 14, Appl
666	4	19.0	89	2	US-08-705-625-2	Sequence 2, Appli
667	4	19.0	89	2	US-08-422-333-13	Sequence 13, Appl
668	4	19.0	89	4	US-09-203-623-3	Sequence 3, Appli
669	4	19.0	89	2	US-09-220-574-2	Sequence 2, Appli
670	4	19.0	89	4	US-08-905-223-433	Sequence 433, App
671	4	19.0	89	6	5187153-21	Patent NO. 5187153
672	4	19.0	89	6	5223482-23	Patent NO. 5223482
673	4	19.0	89	6	5466783-3	Patent NO. 5466783
674	4	19.0	90	4	5220013-24	Patent NO. 5220013
675	4	19.0	92	2	US-09-334-601-29	Sequence 29, Appl
676	4	19.0	92	4	US-09-334-601-30	Sequence 30, Appl
677	4	19.0	93	1	US-09-247-155-114	Sequence 114, App
678	4	19.0	94	1	US-08-266-451B-20	Sequence 20, Appl
679	4	19.0	94	2	US-08-748-725-20	Sequence 20, Appl
680	4	19.0	94	4	US-08-488-551B-640	Sequence 640, App
681	4	19.0	96	4	US-08-465-343A-11	Sequence 11, Appl
682	4	19.0	98	1	US-08-118-270-337	Sequence 337, App
683	4	19.0	98	2	US-09-010-928B-20	Sequence 20, Appl
684	4	19.0	98	4	US-09-214-095D-4	Sequence 4, Appli

No.			Patent/Sequence	n	Sequence
685	19.0	4	PCT-US93-08528-337	5	Sequence 337, App
686	19.0	4	US-08-611-510-7	5	Sequence 7, Appl
687	19.0	4	US-08-672-345C-91	2	Sequence 91, Appl
688	19.0	4	US-08-894-699-34	3	Sequence 34, Appl
689	19.0	4	US-09-444-410-34	4	Sequence 34, Appl
690	19.0	4	US-09-214-095D-88	4	Sequence 88, Appl
691	19.0	4	US-09-214-095D-92	4	Sequence 92, Appl
692	19.0	4	US-09-127-424-2	3	Sequence 2, Appli
693	19.0	4	US-09-214-095D-96	4	Sequence 96, Appl
694	19.0	4	US-09-495-082-2	4	Sequence 2, Appli
695	19.0	4	US-09-461-697-441	4	Sequence 441, App
696	19.0	4	US-09-247-155-111	4	Sequence 111, App
697	19.0	4	US-08-936-165A-432	4	Sequence 432, App
698	19.0	4	US-07-998-003A-99	1	Sequence 99, Appl
699	19.0	4	US-08-453-274B-99	1	Sequence 99, Appl
700	19.0	4	US-08-241-853-11	1	Sequence 11, Appl
701	19.0	4	US-08-326-117B-4	1	Sequence 4, Appli
702	19.0	4	US-08-453-695A-99	1	Sequence 99, Appl
703	19.0	4	US-08-268-161A-99	1	Sequence 99, Appl
704	19.0	4	US-08-850-917-11	2	Sequence 11, Appl
705	19.0	4	US-08-301-915-2	2	Sequence 2, Appli
706	19.0	4	US-08-464-517-28	2	Sequence 28, Appl
707	19.0	4	US-08-453-702A-99	2	Sequence 99, Appl
708	19.0	4	US-08-246-361A-28	3	Sequence 28, Appl
709	19.0	4	US-08-982-129-4	3	Sequence 4, Appli
710	19.0	4	US-08-488-551B-639	3	Sequence 639, App
711	19.0	4	US-08-524-694A-2	3	Sequence 2, Appli
712	19.0	4	US-08-463-772-28	3	Sequence 28, Appl
713	19.0	4	US-08-867-381A-4	3	Sequence 4, Appli
714	19.0	4	US-09-099-639-99	3	Sequence 99, Appl
715	19.0	4	US-09-521-144-4	4	Sequence 4, Appli
716	19.0	4	PCT-US93-05000-28	5	Sequence 28, Appl
717	19.0	4	PCT-US93-12588-99	5	Sequence 99, Appl
718	19.0	4	PCT-US95-08071-99	5	Sequence 99, Appl
719	19.0	4	US-08-440-354-2	2	Sequence 2, Appli
720	19.0	4	US-08-463-087-2	2	Sequence 2, Appli
721	19.0	4	US-08-820-825-14	2	Sequence 14, Appl
722	19.0	4	US-08-850-910A-30	2	Sequence 30, Appl
723	19.0	4	US-09-307-817-14	4	Sequence 14, Appl
724	19.0	4	US-09-452-817-2	4	Sequence 2, Appli
725	19.0	4	5204252-6	6	Patent No. 5204252
726	19.0	4	5204252-7	6	Patent No. 5204252
727	19.0	4	US-08-409-731A-2	2	Sequence 2, Appli
728	19.0	4	US-08-470-298B-2	2	Sequence 2, Appli
729	19.0	4	US-09-023-073A-2	2	Sequence 2, Appli
730	19.0	4	US-08-388-353-640	3	Sequence 640, App
731	19.0	4	US-09-361-737-2	4	Sequence 2, Appli
732	19.0	4	US-08-936-165A-430	4	Sequence 430, App
733	19.0	4	US-08-436-463-15	1	Sequence 15, Appl
734	19.0	4	US-08-436-463-17	1	Sequence 17, Appl
735	19.0	4	US-08-388-353-639	3	Sequence 639, App
736	19.0	4	US-09-240-274-51	4	Sequence 51, Appl
737	19.0	4	US-09-240-274-52	4	Sequence 52, Appl
738	19.0	4	US-09-240-274-53	4	Sequence 53, Appl
739	19.0	4	US-09-240-274-166	4	Sequence 166, App
740	19.0	4	US-08-436-463-8	1	Sequence 8, Appli
741	19.0	4	US-08-024-253-8	1	Sequence 1, Appli
742	19.0	4	US-08-672-345C-1	2	Sequence 1, Appli
743	19.0	4	US-08-672-345C-3	2	Sequence 3, Appli
744	19.0	4	US-08-672-345C-4	2	Sequence 4, Appli
745	19.0	4	US-09-214-095D-1	4	Sequence 1, Appli
746	19.0	4	US-09-121-179-1	4	Sequence 1, Appli
747	19.0	4	US-09-214-095D-2	4	Sequence 2, Appli
748	19.0	4	US-09-214-095D-3	4	Sequence 3, Appli
749	19.0	4	US-09-214-095D-121	4	Sequence 121, App
750	19.0	4	US-08-122-546-14	1	Sequence 14, Appl
751	19.0	4	US-08-422-613-2	1	Sequence 2, Appli
752	19.0	4	US-08-764-938-14	3	Sequence 14, Appl
753	19.0	4	US-08-131-052-14	3	Sequence 14, Appl
754	19.0	4	US-09-131-053A-14	4	Sequence 14, Appl
755	19.0	4	US-08-945-997C-20	4	Sequence 20, Appl
756	19.0	4	US-08-905-223-336	4	Sequence 336, App

Sequence	Patent	n	grp			No.
Sequence 98, Appl	US-08-545-809A-98	3	111	19.0	4	758
Sequence 2, Appli	US-09-470-614-2	4	111	19.0	4	759
Sequence 72, Appl	US-08-081-539-72	2	111	19.0	4	760
Sequence 72, Appl	US-08-466-647-72	1	112	19.0	4	761
Sequence 4, Appli	US-09-193-877-4	2	112	19.0	4	762
Sequence 75, Appl	US-08-857-076-75	4	112	19.0	4	763
Sequence 6, Appli	US-08-157-370-6	4	112	19.0	4	764
Patent No. 549663	549663-9	6	112	19.0	4	765
Sequence 16, Appl	US-08-436-463-16	1	114	19.0	4	766
Sequence 441, App	US-08-637-759B-441	2	114	19.0	4	767
Sequence 441, App	US-08-871-355A-441	3	114	19.0	4	768
Sequence 2, Appli	US-09-203-623-2	4	114	19.0	4	769
Sequence 441, App	US-09-201-945-441	4	114	19.0	4	770
Sequence 2, Appli	US-08-558-735-2	2	115	19.0	4	771
Sequence 12, Appl	US-08-481-956A-12	2	115	19.0	4	772
Sequence 25, Appl	US-08-580-988-25	2	115	19.0	4	773
Sequence 12, Appl	US-08-629-291A-12	2	115	19.0	4	774
Sequence 12, Appl	US-08-658-335B-12	2	115	19.0	4	775
Sequence 2, Appli	US-08-906-480-2	4	115	19.0	4	776
Patent No. 5215889	5215889-3	6	115	19.0	4	777
Sequence 67, Appl	US-08-478-039-67	1	116	19.0	4	778
Sequence 74, Appl	US-08-478-039-74	1	116	19.0	4	779
Sequence 103, App	US-08-478-039-103	1	116	19.0	4	780
Sequence 67, Appl	US-08-476-349A-67	1	116	19.0	4	781
Sequence 74, Appl	US-08-476-349A-74	1	116	19.0	4	782
Sequence 103, App	US-08-476-349A-103	1	116	19.0	4	783
Sequence 96, Appl	US-09-228-986-96	4	116	19.0	4	784
Sequence 62, Appl	US-08-969-683A-62	4	117	19.0	4	785
Patent No. 5514582	5514582-15	6	117	19.0	4	786
Sequence 18, Appl	US-09-370-253-18	2	118	19.0	4	787
Sequence 10, Appl	US-08-737-560A-10	2	119	19.0	4	788
Sequence 80, Appl	US-08-023-082A-80	3	119	19.0	4	789
Sequence 93, Appl	US-08-545-809A-93	3	119	19.0	4	790
Sequence 397, App	US-08-858-207A-397	4	119	19.0	4	791
Sequence 31, Appl	US-08-293-155A-37	2	120	19.0	4	792
Sequence 32, Appl	US-08-290-592E-31	2	120	19.0	4	793
Sequence 90, Appl	US-08-290-592E-32	2	120	19.0	4	794
Sequence 91, Appl	US-08-553-501A-30	2	120	19.0	4	795
Sequence 2, Appli	US-08-553-501A-91	2	120	19.0	4	796
Sequence 90, Appl	US-08-647-144-2	3	120	19.0	4	797
Sequence 91, Appl	US-09-205-231-90	3	120	19.0	4	798
Sequence 219, App	US-09-205-231-91	3	120	19.0	4	799
Sequence 28, Appl	US-09-199-637A-219	4	120	19.0	4	800
Sequence 29, Appl	US-09-025-769B-23	5	120	19.0	4	801
Sequence 31, Appl	PCT-US95-10053-28	5	120	19.0	4	802
Sequence 32, Appl	PCT-US95-10053-29	5	120	19.0	4	803
Sequence 64, Appl	PCT-US96-09448-31	5	120	19.0	4	804
Sequence 64, Appl	PCT-US96-09448-32	5	120	19.0	4	805
Sequence 37, Appl	US-08-483-636-64	2	121	19.0	4	806
Sequence 61, Appl	US-08-483-632-64	2	121	19.0	4	807
Sequence 1, Appli	US-09-025-769B-37	4	121	19.0	4	808
Sequence 143, App	US-09-025-769B-61	4	121	19.0	4	809
Sequence 17, Appl	US-07-734-534A-1	1	122	19.0	4	810
Sequence 14, Appl	US-08-276-852-143	2	122	19.0	4	811
Sequence 143, App	US-08-487-748A-17	1	122	19.0	4	812
Sequence 3, Appli	US-08-436-463-14	1	122	19.0	4	813
Sequence 143, App	US-08-899-575-143	1	122	19.0	4	814
Sequence 51, Appl	US-08-716-588-3	1	122	19.0	4	815
Sequence 3, Appli	US-08-899-575-143	1	122	19.0	4	816
Sequence 51, Appl	US-08-706-741B-51	2	122	19.0	4	817
Sequence 17, Appl	US-08-966-444-3	2	122	19.0	4	818
Sequence 17, Appl	US-08-924-695A-51	3	122	19.0	4	819
Sequence 17, Appl	US-08-398-633-17	3	122	19.0	4	820
Sequence 17, Appl	US-08-480-070C-17	3	122	19.0	4	821
Sequence 143, App	US-08-609-583A-17	3	122	19.0	4	822
Sequence 159, App	US-08-937-399-17	4	122	19.0	4	823
Sequence 159, App	PCT-US95-08743-143	3	123	19.0	4	824
Sequence 12, Appli	US-08-483-577A-159	3	123	19.0	4	825
Sequence 5, Appli	US-08-897-438-159	2	124	19.0	4	826
Sequence 14, Appli	US-08-743-200-12	2	124	19.0	4	827
	US-08-991-890-5	3	124	19.0	4	828
	US-08-751-359-14	4	124	19.0	4	829

Wed Nov 13 14:19:08 2002

#					Reference	Sequence
831	4	19.0	124	4	US-08-907-146-14	Sequence 14, Appl
832	4	19.0	124	4	US-08-858-207A-498	Sequence 498, App
833	4	19.0	125	1	US-08-478-039-69	Sequence 69, Appl
834	4	19.0	125	1	US-08-476-349A-69	Sequence 69, Appl
835	4	19.0	126	1	US-08-276-852-142	Sequence 142, App
836	4	19.0	126	1	US-08-899-575-142	Sequence 142, App
837	4	19.0	126	1	US-08-899-575-142	Sequence 142, App
838	4	19.0	126	1	US-08-751-359-11	Sequence 11, Appl
839	4	19.0	126	4	US-08-907-146-11	Sequence 11, Appl
840	4	19.0	126	5	PCT-US95-08743-142	Sequence 142, App
841	4	19.0	126	5	PCT-US96-03916-15	Sequence 15, Appl
842	4	19.0	127	3	US-08-637-759B-334	Sequence 334, App
843	4	19.0	127	3	US-08-871-355A-334	Sequence 334, App
844	4	19.0	127	4	US-09-201-945-334	Patent No. 5242807
845	4	19.0	128	2	US-08-820-170A-7	Sequence 7, Appli
846	4	19.0	128	2	US-08-318-837-51	Sequence 51, Appl
847	4	19.0	128	3	US-09-055-699-7	Sequence 7, Appli
848	4	19.0	128	4	US-09-273-565-7	Sequence 7, Appli
849	4	19.0	128	4	US-08-348-548-2	Sequence 2, Appli
850	4	19.0	128	4	US-08-885-469-2	Sequence 2, Appli
851	4	19.0	128	4	US-09-565-538-7	Sequence 7, Appli
852	4	19.0	128	4	US-09-228-986-102	Sequence 102, App
853	4	19.0	128	5	PCT-US95-15716-2	Sequence 2, Appli
854	4	19.0	129	4	US-09-187-049-5	Sequence 2, Appli
855	4	19.0	129	4	US-09-476-482-8	Sequence 8, Appli
856	4	19.0	129	6	5428135-6	Patent No. 5428135
857	4	19.0	130	1	US-08-478-039-66	Sequence 66, Appl
858	4	19.0	130	1	US-08-478-039-106	Sequence 106, App
859	4	19.0	130	1	US-08-476-349A-66	Sequence 66, Appl
860	4	19.0	130	1	US-08-476-349A-106	Sequence 106, App
861	4	19.0	130	2	US-08-630-822A-90	Sequence 90, Appl
862	4	19.0	130	2	US-09-005-069-90	Sequence 90, Appl
863	4	19.0	131	2	US-08-477-451-30	Sequence 30, Appl
864	4	19.0	131	2	US-08-820-825-15	Sequence 15, Appl
865	4	19.0	131	2	US-08-850-910A-39	Sequence 39, Appl
866	4	19.0	131	2	US-08-850-910A-41	Sequence 41, Appl
867	4	19.0	131	2	US-08-850-910A-43	Sequence 43, Appl
868	4	19.0	131	2	US-08-850-910A-46	Sequence 46, Appl
869	4	19.0	131	4	US-09-307-817-15	Sequence 15, Appl
870	4	19.0	132	2	US-08-470-298B-12	Sequence 12, Appl
871	4	19.0	132	2	US-08-850-910A-47	Sequence 47, Appl
872	4	19.0	132	4	US-08-159-106-15	Sequence 15, Appl
873	4	19.0	132	4	US-09-043-646-2	Sequence 13, Appl
874	4	19.0	134	1	US-08-328-152A-13	Sequence 2, Appli
875	4	19.0	135	1	US-08-409-731A-8	Sequence 8, Appli
876	4	19.0	135	2	US-08-336-618-20	Sequence 20, Appl
877	4	19.0	135	2	US-08-470-298B-8	Sequence 8, Appli
878	4	19.0	135	2	US-08-820-825-2	Sequence 2, Appli
879	4	19.0	135	2	US-08-847-724-4	Sequence 4, Appli
880	4	19.0	135	2	US-08-847-724-6	Sequence 6, Appli
881	4	19.0	135	3	US-09-023-073A-8	Sequence 8, Appli
882	4	19.0	135	3	US-08-899-031-1	Sequence 1, Appli
883	4	19.0	135	3	US-08-899-031-3	Sequence 3, Appli
884	4	19.0	137	3	US-08-899-031-4	Sequence 4, Appli
885	4	19.0	137	3	US-09-091-117-1	Sequence 1, Appli
886	4	19.0	135	4	US-09-307-817-2	Sequence 2, Appli
887	4	19.0	135	4	US-09-361-737-8	Sequence 7, Appli
888	4	19.0	136	1	US-08-477-451-31	Sequence 31, Appl
889	4	19.0	136	3	US-08-601-184-7	Sequence 5, Appli
890	4	19.0	136	3	US-09-136-879-5	Sequence 5, Appli
891	4	19.0	137	3	US-09-053-197A-31	Sequence 19, Appl
892	4	19.0	137	3	US-09-456-830-41	Sequence 22, Appl
893	4	19.0	137	4	US-09-085-761A-31	Sequence 10, Appl
894	4	19.0	138	2	US-08-686-878A-14	Sequence 14, Appl
895	4	19.0	138	4	US-09-191-647-4	Sequence 2, Appli
896	4	19.0	138	4	US-09-540-245A-4	Sequence 4, Appli
897	4	19.0	139	2	US-08-553-501A-59	Sequence 15, Appl
898	4	19.0	139	3	US-08-950-720A-15	Sequence 59, Appl
899	4	19.0	139	3	US-08-991-890-2	Sequence 2, Appli
900	4	19.0	139	3	US-09-205-231-59	Sequence 172, App
901	4	19.0	139	4	US-08-444-818-172	
904	4	19.0	139	4	US-08-857-076-13	Sequence 13, Appl
905	4	19.0	139	4	US-08-874-102-36	Sequence 36, Appl
906	4	19.0	139	4	US-08-936-165A-409	Sequence 409, App
907	4	19.0	140	2	US-08-483-636-4	Sequence 4, Appli
908	4	19.0	140	2	US-08-483-632-4	Sequence 4, Appli
909	4	19.0	141	2	US-08-345-321-10	Sequence 10, Appl
910	4	19.0	141	2	US-08-483-636-10	Sequence 10, Appl
911	4	19.0	141	2	US-08-483-636-12	Sequence 12, Appl
912	4	19.0	141	2	US-08-483-632-10	Sequence 10, Appl
913	4	19.0	141	2	US-08-483-632-12	Sequence 12, Appl
914	4	19.0	142	4	US-09-456-830-31	Sequence 31, Appl
915	4	19.0	142	4	US-09-456-830-47	Sequence 47, Appl
916	4	19.0	142	4	US-09-456-830-59	Sequence 59, Appl
917	4	19.0	142	6	5242807-2	Patent No. 5242807
918	4	19.0	143	1	US-08-436-463-2	Sequence 2, Appli
919	4	19.0	143	4	US-09-129-075-12	Sequence 12, Appl
920	4	19.0	143	4	US-08-868-699A-4	Sequence 4, Appli
921	4	19.0	143	4	US-09-757-014-4	Sequence 4, Appli
922	4	19.0	144	4	US-08-311-185-60	Sequence 60, Appl
923	4	19.0	144	3	US-09-082-614A-60	Sequence 60, Appl
924	4	19.0	144	4	US-09-045-973-9	Sequence 9, Appli
925	4	19.0	144	4	US-09-319-989-4	Sequence 4, Appli
926	4	19.0	145	2	US-08-686-599A-20	Sequence 20, Appl
927	4	19.0	147	1	US-08-276-151-4	Sequence 4, Appli
928	4	19.0	147	3	US-08-385-590A-4	Sequence 4, Appli
929	4	19.0	147	3	US-09-021-520-4	Sequence 4, Appli
930	4	19.0	148	1	US-09-388-917-5	Sequence 5, Appli
931	4	19.0	148	1	US-08-233-788A-43	Sequence 43, Appl
932	4	19.0	148	2	US-08-392-625-18	Sequence 18, Appl
933	4	19.0	148	2	US-08-466-961A-18	Sequence 18, Appl
934	4	19.0	148	2	US-08-645-193B-13	Sequence 13, Appl
935	4	19.0	149	4	US-08-808-599A-40	Sequence 40, Appl
936	4	19.0	150	6	5310729-2	Patent No. 5310729
937	4	19.0	151	1	US-08-485-455D-51	Sequence 51, Appl
938	4	19.0	151	2	US-08-722-050-5	Sequence 5, Appli
939	4	19.0	151	2	US-08-482-130C-51	Sequence 51, Appl
940	4	19.0	151	3	US-08-484-211C-51	Sequence 51, Appl
941	4	19.0	151	3	US-08-906-769-51	Sequence 51, Appl
942	4	19.0	151	3	US-08-906-616-51	Sequence 51, Appl
943	4	19.0	151	4	US-08-817-799-51	Sequence 51, Appl
944	4	19.0	151	4	US-08-485-443B-51	Sequence 51, Appl
945	4	19.0	151	4	US-08-639-075A-51	Sequence 51, Appl
946	4	19.0	151	4	US-08-012-431-51	Sequence 51, Appl
947	4	19.0	151	1	US-09-012-692-51	Sequence 51, Appl
948	4	19.0	151	5	US-08-906-613-51	Sequence 51, Appl
949	4	19.0	154	5	PCT-US95-14442A-51	Sequence 26, Appl
950	4	19.0	154	4	US-09-193-104-26	Sequence 110, App
951	4	19.0	154	6	US-09-247-155-110	Patent No. 5183734
952	4	19.0	156	3	5183734-1	Sequence 10, Appl
953	4	19.0	156	3	US-08-826-246-10	Sequence 10, Appl
954	4	19.0	156	3	US-08-944-495-10	Sequence 10, Appl
955	4	19.0	156	4	US-08-126-640-11	Sequence 11, Appl
956	4	19.0	156	4	US-08-925-588-10	Sequence 10, Appl
957	4	19.0	156	4	US-09-056-556-227	Sequence 227, App
958	4	19.0	157	1	US-09-288-292A-11	Sequence 11, Appl
959	4	19.0	157	3	US-08-257-999-4	Sequence 4, Appli
960	4	19.0	157	3	US-08-811-949-41	Sequence 5, Appli
961	4	19.0	158	1	US-08-992-176-5	Sequence 5, Appli
962	4	19.0	158	3	US-07-778-156-3	Sequence 3, Appli
963	4	19.0	158	3	US-08-422-166-3	Sequence 19, Appl
964	4	19.0	158	3	US-09-053-197A-19	Sequence 22, Appl
965	4	19.0	158	4	US-09-215-221-22	Sequence 10, Appl
966	4	19.0	158	4	US-09-126-177-10	Sequence 19, Appl
967	4	19.0	158	4	US-09-085-761A-19	Sequence 2, Appli
968	4	19.0	159	1	US-08-414-926A-2	Sequence 21, Appl
969	4	19.0	159	2	US-08-414-926A-21	Sequence 21, Appl
970	4	19.0	159	2	US-08-926-922-2	Sequence 2, Appli
971	4	19.0	159	2	US-08-926-922-21	Sequence 21, Appl
972	4	19.0	159	3	US-09-253-682-2	Sequence 2, Appli
973	4	19.0	159	3	US-09-253-682-21	Sequence 21, Appl
974	4	19.0	159	3	US-08-991-890-4	Sequence 4, Appli
975	4	19.0	159	3	US-09-527-657-2	Sequence 2, Appli
976	4	19.0	159	4	US-09-527-657-21	Sequence 21, Appl

977	4	19.0	US-08-014-153D-8	1	160	Sequence 8, Appli
978	4	19.0	US-08-883-070-7	2	160	Sequence 7, Appli
979	4	19.0	US-08-997-080-44	2	161	Sequence 44, Appl
980	4	19.0	US-08-997-080-70	2	161	Sequence 70, Appl
981	4	19.0	US-08-997-362-44	2	161	Sequence 44, Appl
982	4	19.0	US-08-997-362-70	2	161	Sequence 70, Appl
983	4	19.0	US-08-873-970-44	3	161	Sequence 44, Appl
984	4	19.0	US-08-873-970-70	4	161	Sequence 70, Appl
985	4	19.0	US-09-095-855-44	4	161	Sequence 44, Appl
986	4	19.0	US-09-095-855-70	4	161	Sequence 70, Appl
987	4	19.0	US-08-705-347A-44	4	161	Sequence 44, Appl
988	4	19.0	US-09-324-542-44	4	161	Sequence 44, Appl
989	4	19.0	US-09-324-542-70	4	161	Sequence 70, Appl
990	4	19.0	US-08-362-453-15	1	162	Sequence 15, Appl
991	4	19.0	US-08-800-682-7	4	162	Sequence 7, Appli
992	4	19.0	US-08-469-667-11	4	163	Sequence 11, Appl
993	4	19.0	US-09-224-110-11	4	163	Sequence 11, Appl
994	4	19.0	PCT-US95-07289-11	5	163	Sequence 11, Appl
995	4	19.0	US-08-357-125-4	4	164	Sequence 4, Appli
996	4	19.0	US-08-318-193-77	1	164	Sequence 77, Appl
997	4	19.0	US-08-024-330-1	1	165	Sequence 1, Appli
998	4	19.0	US-07-952-840-1	1	165	Sequence 1, Appli
999	4	19.0	US-08-145-995A-11	1	165	Sequence 11, Appl
1000	4	19.0	PCT-US95-03866-5	5	165	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-482-918-56
; Sequence 56, Application US/08482918
; Patent No. 6207417
; GENERAL INFORMATION:
; APPLICANT: Zsebo, Krisztina M.
; APPLICANT: Bosselman, Robert A.
; APPLICANT: Suggs, Sidney V.
; APPLICANT: Martin, Francis H.
; TITLE OF INVENTION: Stem Cell Factor
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,918
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 01017/33005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-482-918-56

Query Match   28.6%; Score 6; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   7 FFRLFN 12
       ||||||
Db 139 FFRLFN 144

RESULT 2
US-09-224-681-56
; Sequence 56, Application US/09224681
; Patent No. 6207454
; GENERAL INFORMATION:
; APPLICANT: Zsebo, Krisztina M.
; APPLICANT: Bosselman, Robert A.
; APPLICANT: Suggs, Sidney V.
; APPLICANT: Martin, Francis H.
; TITLE OF INVENTION: Method for Enhancing the Efficiency of Gene
; TITLE OF INVENTION: Transfer with Stem Cell Factor (SCF) Polypeptide
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,681
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/005,893
; FILING DATE: 12-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/449,653
; FILING DATE: 24-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/982,255
; FILING DATE: 25-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/589,701
; FILING DATE: 01-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/573,616
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/537,198
; FILING DATE: 11-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/422,383
; FILING DATE: 16-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 01017/35199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX:
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```
;
   TOPOLOGY: linear
;  MOLECULE TYPE: Protein
US-09-224-681-56

Query Match      28.6%;   Score 6;  DB 4;  Length 282;
Best Local Similarity 100.0%;  Pred. No. 34;
Matches   6; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    7 FFRLFN 12
        |||||
Db  139 FFRLFN 144

RESULT 3
US-08-336-728A-56
; Sequence 56, Application US/08336728A
; Patent No. 6207802
; GENERAL INFORMATION:
;   APPLICANT: Zsebo, Krisztina M.
;   APPLICANT: Bosselman, Robert A.
;   APPLICANT: Suggs, Sidney V.
;   APPLICANT: Martin, Francis H.
;   TITLE OF INVENTION: Stem Cell Factor
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;     STREET: 6300 Sears Tower, 233 South Wacker Drive
;     CITY: Chicago
;     STATE: Illinois
;     COUNTRY: United States of America
;     ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/336,728A
;     FILING DATE: 09-NOV-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/982,255
;     FILING DATE: 25-NOV-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/589,701
;     FILING DATE: 01-OCT-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/573,616
;     FILING DATE: 24-AUG-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/537,198
;     FILING DATE: 11-JUN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/422,383
;     FILING DATE: 16-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Clough, David W.
;     REGISTRATION NUMBER: 36,107
;     REFERENCE/DOCKET NUMBER: 01017/32956
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 312/474-6300
;     TELEFAX: 312/474-0448
;     TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 56:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 282 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-728A-56

Query Match      28.6%;  Score 6;  DB 4;  Length 282;
```

```
Best Local Similarity 100.0%;  Pred. No. 34;
Matches   6; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    7 FFRLFN 12
        |||||
Db  139 FFRLFN 144

RESULT 4
US-09-268-992-51
; Sequence 51, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;   APPLICANT: Chen, H.
;   APPLICANT: Freimer, N.
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;   TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;   FILE REFERENCE: 7853-138
;   CURRENT APPLICATION NUMBER: US/09/268,992
;   CURRENT FILING DATE: 1999-03-16
;   EARLIER APPLICATION NUMBER: 09/236,134
;   EARLIER FILING DATE: 1999-01-22
;   EARLIER APPLICATION NUMBER: 60/106,056
;   EARLIER FILING DATE: 1998-10-28
;   EARLIER APPLICATION NUMBER: 60/088,312
;   EARLIER FILING DATE: 1998-06-05
;   EARLIER APPLICATION NUMBER: 60/078,044
;   EARLIER FILING DATE: 1998-03-16
;   NUMBER OF SEQ ID NOS: 84
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 51
;   LENGTH: 446
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-268-992-51

Query Match      28.6%;  Score 6;  DB 4;  Length 446;
Best Local Similarity 100.0%;  Pred. No. 51;
Matches   6; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   10 LFNRSF 15
        |||||
Db  174 LFNRSF 179

RESULT 5
US-09-268-992-64
; Sequence 64, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
;   APPLICANT: Chen, H.
;   APPLICANT: Freimer, N.
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
;   TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
;   FILE REFERENCE: 7853-138
;   CURRENT APPLICATION NUMBER: US/09/268,992
;   CURRENT FILING DATE: 1999-03-16
;   EARLIER APPLICATION NUMBER: 09/236,134
;   EARLIER FILING DATE: 1999-01-22
;   EARLIER APPLICATION NUMBER: 60/106,056
;   EARLIER FILING DATE: 1998-10-28
;   EARLIER APPLICATION NUMBER: 60/088,312
;   EARLIER FILING DATE: 1998-06-05
;   EARLIER APPLICATION NUMBER: 60/078,044
;   EARLIER FILING DATE: 1998-03-16
;   NUMBER OF SEQ ID NOS: 84
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
;   LENGTH: 466
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-268-992-64
```

Query Match Best Local Similarity 28.6%; Score 6; DB 4; Length 466;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   10 LFNRSF 15
        ||||||
Db 194 LFNRSF 199

RESULT 6
US-09-268-992-4
; Sequence 4, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-4

Query Match Best Local Similarity 28.6%; Score 6; DB 4; Length 477;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   10 LFNRSF 15
        |||||
Db 225 LFNRSF 230

RESULT 7
US-09-268-992-2
; Sequence 2, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-2

Query Match Best Local Similarity 28.6%; Score 6; DB 4; Length 495;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   10 LFNRSF 15
        ||||||
Db 223 LFNRSF 228

RESULT 8
US-07-952-853-6
; Sequence 6, Application US/07952853
; Patent No. 5863783
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flipphi, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES
; TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,853
; FILING DATE: 19921125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-952-853-6

Query Match Best Local Similarity 28.6%; Score 6; DB 2; Length 499;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 SGSLST 6
        ||||||
Db 264 SGSLST 269

RESULT 9
US-08-914-848-6
; Sequence 6, Application US/08914848
; Patent No. 5989887

GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaf, Leendert H.
; APPLICANT: Flipphi, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: MORRISON & FOERSTER
;   STREET: 755 Page Mill Road
;   CITY: Palo Alto
;   STATE: California
;   ZIP: 94304-1018
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/914,848
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/952,853
;   FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mirashige, Kate H.
;   REGISTRATION NUMBER: 29,959
;   REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-813-5600
;   TELEFAX: 415-494-0792
;   TELEX: 706141
US-08-914-848-6
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 499 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein

Query Match   28.6%; Score 6; DB 2; Length 499;
Best Local Similarity   100.0%; Pred. No. 57;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
         ||||||
Db   264 SGSLST 269

RESULT 10
US-09-014-969-11
; Sequence 11, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
;   APPLICANT: Jacobs, Kenneth
;   APPLICANT: McCoy, John M.
;   APPLICANT: LaVallie, Edward R.
;   APPLICANT: Racie, Lisa A.
;   APPLICANT: Merberg, David
;   APPLICANT: Treacy, Maurice
;   APPLICANT: Spaulding, Vikki
;   APPLICANT: Agostino, Michael J.
;   TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;   TITLE OF INVENTION: ENCODING THEM
;   NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genetics Institute, Inc.
;   STREET: 87 CambridgePark Drive
;   CITY: Cambridge
;   STATE: MA
;   COUNTRY: U.S.A.
;   ZIP: 02140
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/014,969
;   FILING DATE:
;   CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Sprunger, Suzanne A.
;   REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 498-8284
;   TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 635 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-014-969-11

Query Match   28.6%; Score 6; DB 2; Length 635;
Best Local Similarity   100.0%; Pred. No. 70;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
         ||||||
Db   361 SGSLST 366

RESULT 11
US-07-745-206A-7
; Sequence 7, Application US/07745206A
; Patent No. 5429921
; GENERAL INFORMATION:
;   APPLICANT: Harpold, Michael
;   APPLICANT: Ellis, Steven
;   APPLICANT: Williams, Mark
;   APPLICANT: McCue, Ann
;   APPLICANT: Feldman, Daniel
;   TITLE OF INVENTION: Human Calcium Channel Compositions and
;   TITLE OF INVENTION: Methods
;   NUMBER OF SEQUENCES: 32
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fitch, Even, Tabin & Flannery
;   STREET: 135 S. LaSalle
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: U.S.A.
;   ZIP: 60603
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/745,206A
;   FILING DATE: 19910815
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Feder, Scott B
;   REFERENCE/DOCKET NUMBER: 51504
; TELECOMMUNICATION INFORMATION:

;   TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1968 amino acids
;   TYPE: AMINO ACID
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
US-07-745-206A-7

Query Match   28.6%;   Score 6;   DB 1;   Length 1968;
Best Local Similarity 100.0%;   Pred. No. 1.9e+02;
Matches   6;   Conservative   0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   6 TFFRLF 11
          ||||||
Db   1351 TFFRLF 1356

RESULT 12
US-08-455-543A-45
;   Sequence 45, Application US/08455543A
;   Patent No. 5792846
;   GENERAL INFORMATION:
;   APPLICANT: Harpold, Michael
;   APPLICANT: Ellis, Steven
;   APPLICANT: Williams, Mark
;   APPLICANT: Feldman, Daniel
;   APPLICANT: McCue, Ann
;   APPLICANT: Brenner, Robert
;   TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;   TITLE OF INVENTION: METHODS
;   NUMBER OF SEQUENCES: 57
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Martin, Haller & McClain
;   STREET: 1660 Union Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101-2926
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/455,543A
;   FILING DATE: May 31, 1995
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/223,305
;   FILING DATE: April 4, 1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/868,354
;   FILING DATE: April 10, 1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/745,206
;   FILING DATE: 15-AUG-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/620,250
;   FILING DATE: 30-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/482,384
;   FILING DATE: 20-FEB-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/603,751
;   FILING DATE: 04-APR-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/US89/01408
;   FILING DATE: 04-APR-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/176,899
;   FILING DATE: 04-APR-1988
;   PRIOR APPLICATION DATA:
;   ATTORNEY/AGENT INFORMATION:

;   NAME: Seidman, Stephanie L.
;   REGISTRATION NUMBER: 33,779
;   REFERENCE/DOCKET NUMBER: 6362-52517
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619)238-0999
;   TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 45:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1968 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: internal
US-08-455-543A-45

Query Match   28.6%;   Score 6;   DB 1;   Length 1968;
Best Local Similarity 100.0%;   Pred. No. 1.9e+02;
Matches   6;   Conservative   0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   6 TFFRLF 11
          ||||||
Db   1351 TFFRLF 1356

RESULT 13
US-08-223-305C-45
;   Sequence 45, Application US/08223305C
;   Patent No. 5851824
;   GENERAL INFORMATION:
;   APPLICANT: Harpold, Michael
;   APPLICANT: Ellis, Steven
;   APPLICANT: Williams, Mark
;   APPLICANT: Feldman, Daniel
;   APPLICANT: McCue, Ann
;   APPLICANT: Brenner, Robert
;   TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;   TITLE OF INVENTION: METHODS
;   NUMBER OF SEQUENCES: 57
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Martin, Haller & McClain
;   STREET: 1660 Union Street
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92101-2926
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/223,305C
;   FILING DATE: April 4, 1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/868,354
;   FILING DATE: April 10, 1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/745,206
;   FILING DATE: 15-AUG-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/620,250
;   FILING DATE: 30-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/482,384
;   FILING DATE: 20-FEB-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/603,751
;   FILING DATE: 04-APR-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/US89/01408
;   FILING DATE: 04-APR-1989
;   PRIOR APPLICATION DATA:

```
       APPLICATION NUMBER: US 07/176,899
       FILING DATE: 04-APR-1988
     ATTORNEY/AGENT INFORMATION:
       NAME: Seidman, Stephanie L.
       REGISTRATION NUMBER: 33,779
       REFERENCE/DOCKET NUMBER: 52516  (P519739)
     TELECOMMUNICATION INFORMATION:
       TELEPHONE: (619)238-0999
       TELEFAX: (619)238-0062
   INFORMATION FOR SEQ ID NO: 45:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 1968 amino acids
       TYPE: amino acid
       STRANDEDNESS: single
       TOPOLOGY: linear
     MOLECULE TYPE: protein
       FRAGMENT TYPE: internal
US-08-223-305C-45

Query Match          28.6%;   Score 6;   DB 2;  Length 1968;
Best Local Similarity   100.0%;   Pred. No. 1.9e+02;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db  1351 TFFRLF 1356

RESULT 14
US-08-311-363-7
; Sequence 7, Application US/08311363
; Patent No. 5876958
; GENERAL INFORMATION:
       APPLICANT: Harpold, Michael
       APPLICANT: Ellis, Steven
       APPLICANT: Williams, Mark
       APPLICANT: Feldman, Daniel
       APPLICANT: McCue, Ann
       APPLICANT: Brenner, Robert
       TITLE OF INVENTION: Human Calcium Channel Compositions and
       TITLE OF INVENTION: Methods
       NUMBER OF SEQUENCES: 32
     CORRESPONDENCE ADDRESS:
       ADDRESSEE: Brown, Martin, Haller & McClain
       STREET: 1660 Union Street
       CITY: San Diego
       STATE: California
       COUNTRY: USA
       ZIP: 92101-2926
     COMPUTER READABLE FORM:
       MEDIUM TYPE: Floppy disk
       COMPUTER: IBM PC compatible
       OPERATING SYSTEM: PC-DOS/MS-DOS
       SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/08/311,363
       FILING DATE:
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER: US 07/745,206
       FILING DATE: 15-AUG-1991
     ATTORNEY/AGENT INFORMATION:
       NAME: Seidman, Stephanie L.
       REGISTRATION NUMBER: 33,779
     TELECOMMUNICATION INFORMATION:
       REFERENCE/DOCKET NUMBER: 6362-51506
       TELEPHONE: (619)238-0999
       TELEFAX: (619)238-0062
   INFORMATION FOR SEQ ID NO: 7:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 1968 amino acids
       TYPE: amino acid
       STRANDEDNESS: unknown
       TOPOLOGY: unknown
```

```
     MOLECULE TYPE: peptide
US-08-311-363-7

Query Match          28.6%;   Score 6;   DB 2;  Length 1968;
Best Local Similarity   100.0%;   Pred. No. 1.9e+02;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db  1351 TFFRLF 1356

RESULT 15
US-07-745-206A-2
; Sequence 2, Application US/07745206A
; Patent No. 5429921
; GENERAL INFORMATION:
       APPLICANT: Harpold, Michael
       APPLICANT: Ellis, Steven
       APPLICANT: Williams, Mark
       APPLICANT: McCue, Ann
       APPLICANT: Feldman, Daniel
       TITLE OF INVENTION: Human Calcium Channel Compositions and
       TITLE OF INVENTION: Methods
       NUMBER OF SEQUENCES: 32
     CORRESPONDENCE ADDRESS:
       ADDRESSEE: Fitch, Even, Tabin & Flannery
       STREET: 135 S. LaSalle
       CITY: Chicago
       STATE: Illinois
       COUNTRY: U.S.A.
       ZIP: 60603
     COMPUTER READABLE FORM:
       MEDIUM TYPE: Floppy disk
       COMPUTER: IBM PC compatible
       OPERATING SYSTEM: PC-DOS/MS-DOS
       SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/07/745,206A
       FILING DATE: 19910815
       CLASSIFICATION: 435
     ATTORNEY/AGENT INFORMATION:
       NAME: Feder, Scott B
       REFERENCE/DOCKET NUMBER: 51504
     TELECOMMUNICATION INFORMATION:
       TELEPHONE: 312-372-7842
   INFORMATION FOR SEQ ID NO: 2:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 2161 amino acids
       TYPE: AMINO ACID
       TOPOLOGY: linear
     MOLECULE TYPE: protein
US-07-745-206A-2

Query Match          28.6%;   Score 6;   DB 1;  Length 2161;
Best Local Similarity   100.0%;   Pred. No. 2e+02;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 TFFRLF 11
         ||||||
Db  1316 TFFRLF 1321

Search completed: November  5, 2002, 11:07:06
Job time : 16.8955 secs
```

```
OM protein - protein search, using sw model
                    GenCore version 5.1.3
              Copyright (c) 1993 - 2002  Compugen Ltd.

Run on:    November  5, 2002, 11:02:02 ; Search time 8.14925 Seconds
                              (without alignments)
                              247.615 Million cell updates/sec

Title:        US-09-833-017B-4
Perfect score:  21
Sequence:     1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table:  OLIGO
              Gapop 60.0 , Gapext 60.0

Searched:     283138 seqs, 96089334 residues

Word size :   0

Total number of hits satisfying chosen parameters:      283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :    PIR_71:*
          1: pir1:*
          2: pir2:*
          3: pir3:*
          4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	33.3	230	2	S46737	hypothetical prote
2	7	33.3	253	2	S70367	stem cell factor s
3	7	33.3	287	2	S70366	stem cell factor 1
4	7	33.3	472	2	AD2284	hypothetical prote
5	7	33.3	565	2	AI0479	probable membrane
6	6	28.6	43	2	A97416	hypothetical prote
7	6	28.6	69	2	T18094	hypothetical prote
8	6	28.6	122	2	B75387	hypothetical prote
9	6	28.6	128	2	AE1330	hypothetical prote
10	6	28.6	134	2	T17655	hypothetical prote
11	6	28.6	155	2	B70370	hypothetical prote
12	6	28.6	181	2	S78572	EST3 protein - yea
13	6	28.6	186	2	A56956	GTP-binding protei
14	6	28.6	197	2	A44363	voltage-gated dihy
15	6	28.6	202	2	H81731	CDP-diacylglycerol
16	6	28.6	213	2	E71511	hypothetical prote
17	6	28.6	214	2	G70348	conserved hypothet
18	6	28.6	271	2	G84030	hemA concentration
19	6	28.6	283	2	C23660	calcium channel pr
20	6	28.6	285	2	E83795	ABC transporter (A
21	6	28.6	287	2	JN0637	stem cell factor p
22	6	28.6	287	2	D97241	phosphohydrolase f
23	6	28.6	288	2	D82578	conserved hypothet
24	6	28.6	289	2	T28922	hypothetical prote
25	6	28.6	294	2	A23660	calcium channel pr
26	6	28.6	294	2	B23660	calcium channel pr
27	6	28.6	295	2	S05005	calcium channel al
28	6	28.6	295	2	F98205	probable transcrip
29	6	28.6	295	2	AC3081	transcription regu
30	6	28.6	296	2	A36366	enhancer-binding p
31	6	28.6	297	2	A35914	transcription fact
32	6	28.6	300	2	E83388	probable transcrip
33	6	28.6	314	1	T45010	probable site-spec
34	6	28.6	329	2	D71316	conserved hypothet
35	6	28.6	329	2	I64160	hypothetical prote
36	6	28.6	340	2	T42995	hypothetical prote
37	6	28.6	345	2	S12788	transcription fact
38	6	28.6	360	2	B86531	Yop translocation
39	6	28.6	360	2	C72091	type III secretion
40	6	28.6	370	2	T48633	hypothetical prote
41	6	28.6	385	2	AF2427	probable glycosyl
42	6	28.6	405	2	A85832	probable colanic a
43	6	28.6	405	2	H64971	probable glycosyl
44	6	28.6	405	2	F90986	probable protein [
45	6	28.6	421	2	C95880	hypothetical prote
46	6	28.6	428	2	T31746	probable glycosyl
47	6	28.6	460	2	AG0965	probable sterol o-
48	6	28.6	472	2	T41684	alpha-L-arabinofur
49	6	28.6	499	2	S39113	hypothetical prote
50	6	28.6	519	2	T24243	hypothetical 62.1
51	6	28.6	561	1	E64980	probable 2-compone
52	6	28.6	561	2	A98996	probable 2-compone
53	6	28.6	561	2	F85841	probable two-compo
54	6	28.6	561	2	AE0777	protein B0212.2 [i
55	6	28.6	575	2	E88651	probable cytochrom
56	6	28.6	583	2	T15644	protein R19.7 [im
57	6	28.6	583	2	G87719	hypothetical prote
58	6	28.6	612	2	T50226	hypothetical prote
59	6	28.6	612	2	T29761	methylmalonyl-CoA
60	6	28.6	638	2	S04640	DNA-directed RNA p
61	6	28.6	649	2	H90163	hypothetical prote
62	6	28.6	671	2	T33491	hypothetical prote
63	6	28.6	699	2	T15920	hypothetical prote
64	6	28.6	726	2	A87334	ABC transporter, H
65	6	28.6	759	2	S53922	PMT6 protein - yea
66	6	28.6	804	2	T37821	probable dna repai
67	6	28.6	841	2	A90669	probable enzyme [1
68	6	28.6	841	2	C64755	yagX protein - Esc
69	6	28.6	841	2	D85519	probable enzyme ya
70	6	28.6	970	2	G84939	hypothetical prote
71	6	28.6	1054	2	D96519	mysoin-like protei
72	6	28.6	1075	2	T38253	hypothetical prote
73	6	28.6	1126	2	S04716	DNA-directed RNA p
74	6	28.6	1184	2	A96638	hypothetical prote
75	6	28.6	1291	2	T17242	hypothetical prote
76	6	28.6	1331	2	S05011	calcium channel al
77	6	28.6	1559	2	T30535	calcium channel al
78	6	28.6	1610	2	A46227	voltage-dependent
79	6	28.6	1646	2	JH0422	voltage-dependent
80	6	28.6	1748	2	JN0786	integrin beta-4 ch
81	6	28.6	1787	2	G81684	excinuclease ABC c
82	6	28.6	1807	2	JC6319	integrin-beta-4 ch
83	6	28.6	1852	2	A37860	calcium channel pr
84	6	28.6	1946	2	AC2141	serine/threonine k
85	6	28.6	2100	2	T38128	t7123.15 protein -
86	6	28.6	2139	2	A44467	voltage-dependent
87	6	28.6	2143	2	JH0427	voltage-dependent
88	6	28.6	2161	2	JH0564	calcium channel al
89	6	28.6	2166	2	S11339	calcium channel pr
90	6	28.6	2171	2	S05054	calcium channel al
91	6	28.6	2181	2	A38198	voltage-dependent
92	6	28.6	2203	2	T42742	calcium channel pr
93	6	28.6	2220	2	A45290	calcium channel pr
94	6	28.6	2271	2	F90073	hypothetical prote
95	6	28.6	3194	2	D71917	toxin-like outer m
96	6	28.6	3512	2	T17121	CPY protein - midg
97	5	28.6	3972	2	S75251	hypothetical prote
98	5	23.8	36	2	B44400	myosin heavy chain
99	5	23.8	42	2	A31918	cathepsin D (EC 3.
100	5	23.8	42	2	JE0005	actin alpha, smoot
101	5	23.8	47	2	D81117	hypothetical prote
102	5	23.8	52	2	D82682	hypothetical prote

No.	%	col	Accession	n	pos	Description
103	23.8	5	F82662	2	53	hypothetical prote
104	23.8	5	AB1896	2	53	hypothetical prote
105	23.8	5	I46468	2	54	actin - sheep (fra
106	23.8	5	T29601	2	66	probable transposa
107	23.8	5	I46473	2	67	alpha-actin - rabb
108	23.8	5	E81042	2	67	hypothetical prote
109	23.8	5	E69447	2	67	hypothetical prote
110	23.8	5	T07256	2	76	hypothetical prote
111	23.8	5	B75271	2	77	hypothetical prote
112	23.8	5	T14995	2	77	translation initia
113	23.8	5	S17872	2	78	hypothetical prote
114	23.8	5	H82705	2	79	hypothetical prote
115	23.8	5	E71430	2	80	hypothetical prote
116	23.8	5	T18102	2	82	hypothetical prote
117	23.8	5	S53038	2	83	hypothetical prote
118	23.8	5	E90475	2	84	hypothetical prote
119	23.8	5	T26333	2	85	flavocytochrome c,
120	23.8	5	A36437	1	87	hypothetical prote
121	23.8	5	E83141	2	89	hypothetical prote
122	23.8	5	A54937	2	90	brain type III sod
123	23.8	5	B45837	2	92	Ig gamma-1 chain C
124	23.8	5	S57260	2	92	nitrite reductase
125	23.8	5	H42518	2	94	Al5L 1lK protein -
126	23.8	5	G36849	2	94	Al5L protein - var
127	23.8	5	T28557	2	94	hypothetical prote
128	23.8	5	E72165	2	94	Al6L protein - var
129	23.8	5	C69449	2	95	conserved hypothet
130	23.8	5	S26652	2	97	Ig gamma-1 chain C
131	23.8	5	E95064	2	97	conserved hypothet
132	23.8	5	G64454	2	97	hypothetical prote
133	23.8	5	G97931	2	97	conserved hypothet
134	23.8	5	AH0965	2	98	probable secreted
135	23.8	5	T42307	2	99	hypothetical prote
136	23.8	5	AC0152	2	99	hypothetical prote
137	23.8	5	S14851	2	100	actin - fruit fly
138	23.8	5	A87463	2	100	conserved hypothet
139	23.8	5	I38611	2	101	calcium channel, v
140	23.8	5	H81352	2	101	small hydrophobic
141	23.8	5	S51974	2	102	probable membrane
142	23.8	5	A86584	2	103	CT631 hypothetical
143	23.8	5	S53942	2	104	probable membrane
144	23.8	5	S51479	2	104	drought-induced pr
145	23.8	5	AG3518	2	104	hypothetical prote
146	23.8	5	E81963	2	105	hypothetical IS110
147	23.8	5	H32537	2	106	T-cell receptor be
148	23.8	5	G97841	2	107	acetyl-CoA acetylt
149	23.8	5	D71114	2	107	hypothetical prote
150	23.8	5	JC1127	2	107	major allergen cha
151	23.8	5	C56413	2	109	hypothetical prote
152	23.8	5	T35249	2	109	gas-vesicle operon
153	23.8	5	JQ1128	2	113	gas-vesicle operon
154	23.8	5	T08234	2	113	transcription fact
155	23.8	5	I51025	2	113	hypothetical prote
156	23.8	5	F81370	2	113	hypothetical prote
157	23.8	5	F71141	2	114	hypothetical prote
158	23.8	5	A59316	1	117	ghrelin precursor
159	23.8	5	S28124	2	117	gas-vesicle operon
160	23.8	5	T49511	2	117	hypothetical prote
161	23.8	5	T30509	2	118	gas vesicle protei
162	23.8	5	S28133	2	119	hypothetical prote
163	23.8	5	H70437	2	119	hypothetical prote
164	23.8	5	H84973	2	120	actin, skeletal mu
165	23.8	5	F71104	2	120	hypothetical prote
166	23.8	5	B71020	2	120	hypothetical prote
167	23.8	5	G81855	2	122	insertion element
168	23.8	5	F83507	2	123	hypothetical prote
169	23.8	5	A26337	2	125	actin, skeletal mu
170	23.8	5	C83138	2	125	hypothetical prote
171	23.8	5	AB2487	2	128	hypothetical prote
172	23.8	5	AE1701	2	128	hypothetical prote
173	23.8	5	B90175	2	129	NADH dehydrogenase
174	23.8	5	PL0098	2	130	Ig heavy chain pre
175	23.8	5	G71212	2	130	hypothetical prote
176	23.8	5	S67869	2	131	probable membrane
177	23.8	5	D69836	2	133	hypothetical prote
178	23.8	5	T43561	2	134	probable transposa
179	23.8	5	S72727	2	134	dedA protein - Myc
180	23.8	5	S38386	2	135	T-cell receptor be
181	23.8	5	A36828	2	136	orf10 protein - Au
182	23.8	5	AG2009	2	137	transcription regu
183	23.8	5	A28258	2	137	actin 5C - fruit f
184	23.8	5	AB2544	2	139	transcription regu
185	23.8	5	A71190	2	140	hypothetical prote
186	23.8	5	S73680	2	140	MG337 homolog P02_
187	23.8	5	H69405	2	141	hypothetical prote
188	23.8	5	S30832	2	143	hypothetical prote
189	23.8	5	JC7309	2	146	agglutinin I - Jer
190	23.8	5	T06471	2	147	core protein - gar
191	23.8	5	A81454	2	148	probable integral
192	23.8	5	F83801	2	148	chorismate mutase
193	23.8	5	E64939	2	148	hypothetical prote
194	23.8	5	B90941	2	148	hypothetical prote
195	23.8	5	F85789	2	149	hypothetical prote
196	23.8	5	D87687	2	151	conserved hypothet
197	23.8	5	C90824	2	151	hypothetical membr
198	23.8	5	MCUR2C	1	151	Spec2c protein - s
199	23.8	5	D87548	2	154	hypothetical prote
200	23.8	5	H82810	2	154	bacterioferritin X
201	23.8	5	H75031	2	154	hypothetical prote
202	23.8	5	D83516	2	155	hypothetical prote
203	23.8	5	A60945	2	155	actin homolog FAT,
204	23.8	5	F71004	2	155	hypothetical prote
205	23.8	5	A71693	2	156	hypothetical prote
206	23.8	5	T16531	2	157	hypothetical prote
207	23.8	5	A71137	2	157	hypothetical prote
208	23.8	5	S75191	2	157	hypothetical prote
209	23.8	5	G97760	2	159	hypothetical prote
210	23.8	5	E95328	2	159	hypothetical prote
211	23.8	5	G84125	2	161	actin - shore pine
212	23.8	5	S26039	2	161	calcineurin regula
213	23.8	5	B71687	2	165	hypothetical prote
214	23.8	5	PS0261	2	166	hypothetical prote
215	23.8	5	AE2392	2	166	probable membrane
216	23.8	5	AG3566	2	167	probable signal pe
217	23.8	5	F97128	2	167	hypothetical prote
218	23.8	5	B81676	2	168	hypothetical prote
219	23.8	5	B71553	2	170	calcineurin regula
220	23.8	5	T25892	2	170	calcineurin regula
221	23.8	5	A33391	1	170	calcineurin regula
222	23.8	5	S34127	1	170	Ig gamma-2a chain
223	23.8	5	JC1220	1	170	calcineurin regula
224	23.8	5	A35944	2	170	hypothetical prote
225	23.8	5	JC7242	2	171	hypothetical prote
226	23.8	5	T28000	2	173	probable transcrip
227	23.8	5	C64161	2	176	conserved hypothet
228	23.8	5	H95399	2	177	calcineurin regula
229	23.8	5	F82786	2	179	conserved hypothet
230	23.8	5	JQ1232	2	180	conserved hypothet
231	23.8	5	B82221	2	181	calcineurin regula
232	23.8	5	AC2826	2	181	hypothetical prote
233	23.8	5	JC1221	2	182	Syd protein [impor
234	23.8	5	C71869	2	183	hypothetical prote
235	23.8	5	AB0127	2	184	hypothetical prote
236	23.8	5	T05925	2	184	lemA protein (U661
237	23.8	5	B97604	2	184	conserved hypothet
238	23.8	5	G97436	2	186	probable membrane
239	23.8	5	AC2655	2	187	probable exported
240	23.8	5	AG0115	2	187	ATP synthase F1, d
241	23.8	5	AD0207	2	188	hypothetical prote
242	23.8	5	H87676	2	189	hypothetical prote
243	23.8	5	D83234	2	184	hypothetical prote
244	23.8	5	F96718	2	186	hypothetical prote
245	23.8	5	AH1870	2	187	hypothetical prote
246	23.8	5	G72748	2	187	conserved hypothet
247	23.8	5	A64332	2	188	probable phosphono
248	23.8	5	E97498	2	189	outer membrane ser

249	23.8	5	190	2	D75401	transcription repr
250	23.8	5	190	2	T26019	hypothetical prote
251	23.8	5	193	1	B69043	conserved hypothet
252	23.8	5	193	2	AF1981	hypothetical prote
253	23.8	5	193	2	AG2389	hypothetical prote
254	23.8	5	194	2	S39829	probable membrane
255	23.8	5	194	2	AI0241	hypothetical prote
256	23.8	5	195	2	S39777	actin beta - pig (
257	23.8	5	195	2	T21997	hypothetical prote
258	23.8	5	195	2	A99346	hypothetical prote
259	23.8	5	196	2	G65039	hypothetical prote
260	23.8	5	196	2	B71232	hypothetical prote
261	23.8	5	197	2	D97061	uncharacterized lo
262	23.8	5	200	2	B69168	hypothetical prote
263	23.8	5	200	2	D95950	probable esterase/
264	23.8	5	202	2	G81897	probable phage ant
265	23.8	5	203	2	G81700	thymidylate kinase
266	23.8	5	204	2	S35256	actin (clone Tac9)
267	23.8	5	206	2	A81231	50S ribosomal prot
268	23.8	5	207	2	D83187	hypothetical prote
269	23.8	5	207	2	AH1190	3-methyladenine DN
270	23.8	5	207	2	AH1548	3-methyladenine DN
271	23.8	5	210	2	H82037	conserved hypothet
272	23.8	5	210	2	E84121	stage II sporulati
273	23.8	5	211	2	T19859	hypothetical prote
274	23.8	5	212	2	E81669	conserved hypothet
275	23.8	5	213	2	S68213	Ig heavy chain (Ma
276	23.8	5	213	2	A61043	actin CA15 - sea s
277	23.8	5	214	2	S17680	fibrinolytic prote
278	23.8	5	214	2	PC4202	monoclonal antibod
279	23.8	5	214	2	C86967	probable membrane
280	23.8	5	214	2	B96755	hypothetical prote
281	23.8	5	215	2	T01349	hypothetical prote
282	23.8	5	216	1	S42716	calcineurin regula
283	23.8	5	216	2	T30657	hypothetical prote
284	23.8	5	216	2	B82612	hypothetical prote
285	23.8	5	216	2	AE2548	hypothetical prote
286	23.8	5	217	2	H71208	conserved hypothet
287	23.8	5	218	1	A70964	probable dedA prot
288	23.8	5	218	2	T23318	hypothetical prote
289	23.8	5	218	2	B72470	hypothetical prote
290	23.8	5	220	2	S68211	Ig heavy chain (Ma
291	23.8	5	221	2	S49220	Ig gamma-1 chain -
292	23.8	5	222	2	E90400	hypothetical prote
293	23.8	5	224	2	H70749	hypothetical prote
294	23.8	5	226	2	S75665	ABC-type transport
295	23.8	5	226	2	T36096	probable secreted
296	23.8	5	228	2	G86647	hypothetical prote
297	23.8	5	230	2	T45555	H+-transporting AT
298	23.8	5	231	2	PC4155	Ig gamma-2b chain
299	23.8	5	231	2	AF0866	DNA mismatch repai
300	23.8	5	231	2	T50232	hypothetical lysin
301	23.8	5	232	2	I46679	alpha-smooth muscl
302	23.8	5	232	2	T29440	probable purine nu
303	23.8	5	233	2	C71288	hypothetical prote
304	23.8	5	233	2	T22977	hypothetical prote
305	23.8	5	233	2	D97118	RNA polymerase sig
306	23.8	5	234	2	E84177	hypothetical prote
307	23.8	5	234	2	G75198	hypothetical prote
308	23.8	5	235	2	A97996	degenerate transpo
309	23.8	5	236	2	I59323	apolipoprotein B m
310	23.8	5	236	2	A87281	hypothetical prote
311	23.8	5	236	2	A75455	hypothetical prote
312	23.8	5	239	1	B1AG55	virB1 protein prec
313	23.8	5	240	2	D82591	protein phosphatas
314	23.8	5	242	2	D87494	conserved hypothet
315	23.8	5	242	2	D84422	60S ribosomal prot
316	23.8	5	243	1	S15388	licheninase (EC 3.
317	23.8	5	246	2	S38950	Ig gamma chain - m
318	23.8	5	247	2	T00692	60S ribosomal prot
319	23.8	5	248	1	F22MG	NADH dehydrogenase
320	23.8	5	248	1	C89472	protein ZC53.3 (im
321	23.8	5	248	2	D86676	transcription regu

hypothetical prote	T30502	2	250	23.8	5	322
hypothetical prote	T45028	2	250	23.8	5	323
hypothetical prote	F83604	2	251	23.8	5	324
probable cobyric a	T15300	2	252	23.8	5	325
conserved hypothet	T31439	2	252	23.8	5	326
Ig heavy chain V r	F87672	2	253	23.8	5	327
protein C28H8.6 [i	B31790	2	254	23.8	5	328
PTS system, IIC co	E88469	2	256	23.8	5	329
hypothetical prote	G95252	2	257	23.8	5	330
hypothetical prote	E98117	2	257	23.8	5	331
probable transcrip	D70341	2	257	23.8	5	332
hypothetical prote	B96003	2	257	23.8	5	333
phosphatidylserine	T32939	2	258	23.8	5	334
hypothetical prote	G83813	2	259	23.8	5	335
hypothetical prote	A70428	2	262	23.8	5	336
nickel transport A	AI3570	2	262	23.8	5	337
hypothetical prote	E72592	2	265	23.8	5	338
hypothetical prote	S76808	2	265	23.8	5	339
hypothetical prote	D70778	2	265	23.8	5	340
conserved hypothet	B90242	2	266	23.8	5	341
ribosomal protein	R5HU7A	2	266	23.8	5	342
ribosomal protein	R5RT7A	1	266	23.8	5	343
carboxylesterase (	A69358	2	266	23.8	5	344
ribosomal protein	A30241	2	266	23.8	5	345
ribosomal protein	T52089	2	266	23.8	5	346
probable sigma fac	I40626	2	266	23.8	5	347
sigma factor E pro	A97109	2	266	23.8	5	348
orotidine-5'-phosp	DEBYOP	1	267	23.8	5	349
urease accessory p	AI3458	1	267	23.8	5	350
interleukin-1 beta	ICHU1B	1	269	23.8	5	351
hypothetical prote	S77394	2	269	23.8	5	352
flagellin [importe	F97101	2	269	23.8	5	353
hypothetical prote	T48780	2	270	23.8	5	354
conserved hypothet	B81752	2	270	23.8	5	355
hypothetical prote	D83542	2	270	23.8	5	356
hypothetical prote	T19761	2	271	23.8	5	357
hypothetical prote	F71490	2	271	23.8	5	358
conserved hypothet	AH3085	2	271	23.8	5	359
hypothetical prote	G87240	2	271	23.8	5	360
probable spermidin	G71296	2	272	23.8	5	361
hypothetical prote	H71441	2	273	23.8	5	362
hypothetical prote	T31528	2	274	23.8	5	363
high-affinity bran	AE3552	2	274	23.8	5	364
hypothetical prote	S74417	2	275	23.8	5	365
protocatechuate 3,	S12962	2	276	23.8	5	366
hypothetical prote	AF2404	2	277	23.8	5	367
probable 2-oxoisov	C69506	2	278	23.8	5	368
hypothetical prote	T26582	2	280	23.8	5	369
conserved hypothet	H95901	2	280	23.8	5	370
hypothetical prote	D72351	2	281	23.8	5	371
hypothetical prote	AF2090	2	282	23.8	5	372
surfactant protein	A48516	2	282	23.8	5	373
3,4-dihydroxypheny	AB0631	2	283	23.8	5	374
hypothetical prote	S61156	2	283	23.8	5	375
probable dimethyls	H64914	2	284	23.8	5	376
hypothetical prote	H90915	2	284	23.8	5	377
probable DMSO redu	E85764	2	284	23.8	5	378
probable DMSO redu	A99201	2	284	23.8	5	379
hypothetical prote	AB0215	2	285	23.8	5	380
3,4-dihydroxypheny	F70348	2	285	23.8	5	381
hypothetical prote	T22624	2	286	23.8	5	382
hypothetical prote	C84567	2	287	23.8	5	383
probable cmaA1 pro	G70974	2	287	23.8	5	384
probable mmaA2 pro	A70614	2	287	23.8	5	385
probable umaA2 pro	B70829	2	287	23.8	5	386
hydroxymethylgluta	AH3492	2	287	23.8	5	387
hypothetical prote	A71148	2	288	23.8	5	388
probable translati	G71304	2	290	23.8	5	389
hypothetical prote	E97637	2	290	23.8	5	390
xanthine dehydroge	AF2860	2	290	23.8	5	391
succinate--CoA lig	B69442	2	291	23.8	5	392
insulin-like growt	I48602	1	291	23.8	5	393
hypothetical prote	T03900	2	291	23.8	5	394

No.	Tm	Len	Accession	AA	Cls	Description
395	23.8	5	T34997	292	2	lysR-type transcri
396	23.8	5	AF3647	292	2	phosphatidylglycer
397	23.8	5	T33825	293	2	hypothetical prote
398	23.8	5	AE1927	293	2	3-ketoacyl-acyl ca
399	23.8	5	E75557	294	2	glutamine cyclotra
400	23.8	5	A97937	294	2	conserved hypothet
401	23.8	5	A95069	294	2	leucine-rich prote
402	23.8	5	H81889	294	2	probable transposa
403	23.8	5	B83433	295	2	translocator prote
404	23.8	5	H91117	295	2	probable transcrip
405	23.8	5	H87216	295	2	Mycolic acid synth
406	23.8	5	G85962	295	2	probable transcrip
407	23.8	5	G83216	297	2	probable transcrip
408	23.8	5	B83823	298	2	endonuclease IV BH
409	23.8	5	AC3185	299	2	transcription regu
410	23.8	5	H64050	300	2	glucose kinase hom
411	23.8	5	B83554	300	1	probable transfera
412	23.8	5	A37766	301	2	SEC14 protein - ye
413	23.8	5	AF2223	301	2	heterodisulfide re
414	23.8	5	H70505	301	2	hypothetical prote
415	23.8	5	C64829	302	2	transcription regu
416	23.8	5	B69903	303	2	aromatic metabolit
417	23.8	5	H97212	303	2	co/Zn/Cd efflux sy
418	23.8	5	C83967	303	2	aspartate carbamoy
419	23.8	5	H95162	305	2	hypothetical prote
420	23.8	5	G98028	305	2	hypothetical prote
421	23.8	5	B64960	305	2	membrane protein y
422	23.8	5	H90965	305	2	hypothetical prote
423	23.8	5	H85813	305	2	hypothetical prote
424	23.8	5	A81966	305	2	probable transcrip
425	23.8	5	H81022	305	2	transcription regu
426	23.8	5	E83200	306	2	probable transcrip
427	23.8	5	AF0315	306	2	conserved hypothet
428	23.8	5	F83254	306	2	probable transcrip
429	23.8	5	S18533	306	2	eryG protein - Sac
430	23.8	5	T19412	306	2	hypothetical prote
431	23.8	5	E82383	307	2	transcription regu
432	23.8	5	T27915	307	2	hypothetical prote
433	23.8	5	E83129	307	2	homoprotocatechuat
434	23.8	5	D97605	307	2	proteinase chain h
435	23.8	5	AF2827	307	2	HFLC protein (impo
436	23.8	5	A03000	308	2	actin 3 - fruit fl
437	23.8	5	AD0502	309	2	homoserine kinase
438	23.8	5	G75286	309	2	serine proteinase
439	23.8	5	AE3558	310	2	transcription regu
440	23.8	5	AG2409	310	2	mRNA-binding prote
441	23.8	5	A24400	311	2	nodulation protein
442	23.8	5	S53836	311	2	ribosomal protein
443	23.8	5	T24625	312	2	hypothetical prote
444	23.8	5	C83989	312	2	two-component resp
445	23.8	5	AD3189	312	2	transcription regu
446	23.8	5	QXBP2L	314	1	hypothetical prote
447	23.8	5	S11787	314	2	nodulation protein
448	23.8	5	AG0485	314	2	2-dehydro-3-deoxyg
449	23.8	5	T29607	314	2	hypothetical prote
450	23.8	5	T18991	314	2	hypothetical prote
451	23.8	5	T29396	315	2	dihydrodipicolinat
452	23.8	5	C95941	315	2	transcription regu
453	23.8	5	B83778	316	2	hypothetical prote
454	23.8	5	H81916	316	2	hypothetical prote
455	23.8	5	B82907	316	2	conserved hypothet
456	23.8	5	T12516	316	2	capsular polysacch
457	23.8	5	F70441	316	2	probable LysR-fami
458	23.8	5	B95373	317	2	hypothetical prote
459	23.8	5	E82520	320	2	hypothetical prote
460	23.8	5	F64384	320	2	dihydrodipicolinat
461	23.8	5	AE3635	320	2	transcription regu
462	23.8	5	E82420	321	2	hypothetical prote
463	23.8	5	T24948	322	2	hypothetical prote
464	23.8	5	T27308	322	2	hypothetical prote
465	23.8	5	D84567	322	2	transcription regu
466	23.8	5	AE3527	322	2	transcription regu
467	23.8	5	G83291	323	2	probable transcrip
468	23.8	5	G1MS	324	1	Ig gamma-1 chain C
469	23.8	5	T37228	324	2	probable sugar tra
470	23.8	5	T48236	324	2	hypothetical prote
471	23.8	5	T48698	325	2	3-dehydroquinate d
472	23.8	5	JC2008	325	2	actin homolog prot
473	23.8	5	T32940	325	2	hypothetical prote
474	23.8	5	D87733	326	2	protein W03D8.9 [1
475	23.8	5	D71521	326	2	probable malate de
476	23.8	5	C81678	326	2	malate dehydrogena
477	23.8	5	S11452	327	2	actin (clone 302)
478	23.8	5	AG3448	327	2	transcription regu
479	23.8	5	JE0375	328	2	carbonic anhydrase
480	23.8	5	C64845	328	2	probable membrane
481	23.8	5	E85651	328	2	hypothetical prote
482	23.8	5	B90791	328	2	hypothetical prote
483	23.8	5	H75432	329	2	hypothetical prote
484	23.8	5	AG1528	329	2	alcohol dehydrogen
485	23.8	5	G2MSA	330	1	Ig gamma-2a chain
486	23.8	5	AE2535	330	2	hypothetical prote
487	23.8	5	S24409	331	2	actin - brown alga
488	23.8	5	AE3372	331	2	toluene tolerance
489	23.8	5	D84752	332	2	hypothetical prote
490	23.8	5	T21279	332	2	hypothetical prote
491	23.8	5	S64337	333	2	transaldolase (EC
492	23.8	5	S75980	333	2	hypothetical prote
493	23.8	5	B72306	333	2	probable threonine
494	23.8	5	B95894	334	2	ADPglyceromanno-he
495	23.8	5	C81952	334	2	ADPglyceromanno-he
496	23.8	5	E81152	334	2	probable iron/asco
497	23.8	5	E83122	334	2	cyclin delta-1 - A
498	23.8	5	S51650	334	2	hypothetical prote
499	23.8	5	AG2467	334	2	Ig gamma-2a chain
500	23.8	5	G2MSAB	335	1	iron-compound ABC
501	23.8	5	D95119	335	2	probable transposa
502	23.8	5	G81877	335	2	probable transposa
503	23.8	5	D81893	335	2	hypothetical prote
504	23.8	5	H97998	336	2	actin - maize (fra
505	23.8	5	T04085	336	2	malate dehydrogena
506	23.8	5	E69852	337	2	hypothetical prote
507	23.8	5	S44207	338	1	flagellar motor sw
508	23.8	5	G83508	338	2	hypothetical prote
509	23.8	5	S45765	338	2	protein M01D1.4 [1
510	23.8	5	C88035	339	2	hypothetical prote
511	23.8	5	T19555	339	2	probable transposi
512	23.8	5	T19600	339	2	hypothetical prote
513	23.8	5	T32676	339	2	probable transposa
514	23.8	5	T19660	339	2	hypothetical prote
515	23.8	5	T28781	339	2	hypothetical prote
516	23.8	5	T15113	339	2	hypothetical prote
517	23.8	5	T24136	339	2	hypothetical prote
518	23.8	5	T29362	339	2	hypothetical prote
519	23.8	5	T20442	339	2	hypothetical prote
520	23.8	5	T15721	339	2	hypothetical prote
521	23.8	5	T18783	339	2	hypothetical prote
522	23.8	5	T26396	339	2	hypothetical prote
523	23.8	5	T22780	339	2	hypothetical prote
524	23.8	5	T23086	339	2	hypothetical prote
525	23.8	5	T19261	339	2	hypothetical prote
526	23.8	5	T27715	339	2	hypothetical prote
527	23.8	5	T25535	339	2	hypothetical prote
528	23.8	5	T25365	339	2	hypothetical prote
529	23.8	5	A35009	340	2	thermopsin (EC 3.4
530	23.8	5	C83124	340	2	probable transcrip
531	23.8	5	P3VVGF	341	2	P3 protein - grape
532	23.8	5	G95888	341	2	hypothetical prote
533	23.8	5	AB1394	341	2	galactosyltransfer
534	23.8	5	AE1769	341	2	galactosyltransfer
535	23.8	5	C69644	342	2	ketol-acid reducto
536	23.8	5	T21098	342	2	protein kinase (EC
537	23.8	5	T51703	342	2	nicotinate-nucleot
538	23.8	5	T17417	343	2	hypothetical prote
539	23.8	5	T17432	343	2	probable transcrip
540	23.8	5	T01432	343	2	secY protein homol
541	23.8	5	KHPGD	344	1	cathepsin D (EC 3.

No.				n	ID	Description
541	23.8	5	344	2	T27410	hypothetical prote
542	23.8	5	345	2	D83551	conserved hypothet
543	23.8	5	347	2	S12955	calcium channel pr
544	23.8	5	347	2	T47070	hypothetical prote
545	23.8	5	347	2	H96504	hypothetical prote
546	23.8	5	347	2	T27315	hypothetical prote
547	23.8	5	348	2	A57234	lin-44 protein pre
548	23.8	5	348	2	T33544	hypothetical prote
549	23.8	5	349	2	T23430	hypothetical prote
550	23.8	5	349	2	B25819	actin, fetal skele
551	23.8	5	350	2	S06758	glycerol-3-phospha
552	23.8	5	350	2	A69834	conserved hypothet
553	23.8	5	350	2	S34557	hypothetical prote
554	23.8	5	350	2	T50581	binding-protein-de
555	23.8	5	350	2	S61646	dolichyl-diphospho
556	23.8	5	350	2	T34002	hypothetical prote
557	23.8	5	350	2	AE1845	hypothetical prote
558	23.8	5	351	2	T40933	hypothetical prote
559	23.8	5	351	2	B70811	hypothetical prote
560	23.8	5	351	2	T51513	hypothetical prote
561	23.8	5	351	2	E97131	N-acetylmuramidase
562	23.8	5	352	2	JS0023	glycerol-3-phospha
563	23.8	5	352	2	A49210	heat shock protein
564	23.8	5	352	2	T27281	hypothetical prote
565	23.8	5	353	1	KIBEMV	thymidine kinase (
566	23.8	5	353	2	S31790	glycerol-3-phospha
567	23.8	5	353	2	S06760	glycerol-3-phospha
568	23.8	5	353	2	E84941	imidazoleglycerol-
569	23.8	5	353	2	AF1606	aminopeptidase P h
570	23.8	5	353	2	B88485	protein F23F12.10
571	23.8	5	355	2	G84040	sulfate ABC transp
572	23.8	5	355	2	S16047	nitrogenase cofact
573	23.8	5	355	2	T20782	hypothetical prote
574	23.8	5	355	2	S77610	probable intercell
575	23.8	5	357	1	ATOQ	actin - Oxyricha
576	23.8	5	357	2	C70805	hypothetical prote
577	23.8	5	358	2	T37805	hypothetical prote
578	23.8	5	360	1	S06759	glycerol-3-phospha
579	23.8	5	360	2	H82081	phospho-N-acetylmu
580	23.8	5	361	2	S65210	hypothetical prote
581	23.8	5	361	1	D69005	hypothetical prote
582	23.8	5	361	2	S68089	actin 2 - Arabidop
583	23.8	5	362	2	S21963	glycerol-3-phospha
584	23.8	5	362	2	A26559	actin type 5, cyto
585	23.8	5	362	2	S68090	actin 8 - Arabidop
586	23.8	5	362	2	G72672	probable integrase
587	23.8	5	363	2	S23137	glycerol-3-phospha
588	23.8	5	363	2	T02522	origin recognition
589	23.8	5	364	2	AG2537	cysteine synthase
590	23.8	5	364	2	D70164	heat shock protein
591	23.8	5	364	2	H72059	2-component sensor
592	23.8	5	364	2	E86563	2-component sensor
593	23.8	5	364	2	AC0589	UDPgalactopyranose
594	23.8	5	365	1	FCECR2	translation releas
595	23.8	5	365	1	A36480	translation releas
596	23.8	5	365	2	A64190	translation releas
597	23.8	5	365	2	AF1952	cysteine synthase
598	23.8	5	365	2	S49007	actin - Pythium ir
599	23.8	5	365	2	A37431	actin, type 1 - Em
600	23.8	5	365	2	C91099	peptide chain rele
601	23.8	5	365	2	G85944	peptide chain rele
602	23.8	5	365	2	AG0871	peptide chain rele
603	23.8	5	366	1	OORTB2	bradykinin recepto
604	23.8	5	366	2	G84849	probable actin [im
605	23.8	5	366	2	AB0109	peptide chain rele
606	23.8	5	366	2	T26624	hypothetical prote
607	23.8	5	367	2	JT0596	actin ArdD - slime
608	23.8	5	367	2	B88969	protein T27B7.1 [i
609	23.8	5	367	2	T46021	hypothetical prote
610	23.8	5	368	2	S71190	heat shock protein
611	23.8	5	369	2	T42295	hypothetical prote
612	23.8	5	369	2	F83250	histidinol-phospha
613	23.8	5	369	2	T22708	hypothetical prote

No.				n	ID	Description
614	23.8	5	370	2	A29664	actin - sea urchin
615	23.8	5	370	2	D70076	ethanolamine trans
616	23.8	5	370	2	F81133	IS1106 transposase
617	23.8	5	370	2	E82369	thiH protein VC006
618	23.8	5	371	2	S77345	hypothetical prote
619	23.8	5	371	2	F90486	dehydrogenase, pro
620	23.8	5	372	2	A75434	prephenate dehydro
621	23.8	5	372	2	B81952	probable type II r
622	23.8	5	373	2	S57262	actin - red alga (
623	23.8	5	373	2	G81302	queuine tRNA-ribos
624	23.8	5	373	2	S54545	hypothetical prote
625	23.8	5	374	1	ATBOB	actin beta - bovin
626	23.8	5	374	2	ATBOG	actin gamma - bovi
627	23.8	5	374	1	JC5818	gamma-actin - huma
628	23.8	5	374	2	A95960	probable cytochrom
629	23.8	5	375	1	ATRB	actin, skeletal mu
630	23.8	5	375	1	ATBOSM	actin, aortic smoo
631	23.8	5	375	1	ATHUB	actin beta - human
632	23.8	5	375	1	ATMSB	actin beta - mouse
633	23.8	5	375	1	ATRTC	actin beta - rat
634	23.8	5	375	1	ATRBB	actin beta, non-mu
635	23.8	5	375	1	ATCHB	actin beta - chick
636	23.8	5	375	1	A48324	actin beta, cytosk
637	23.8	5	375	1	ATHUG	actin gamma 1 - hu
638	23.8	5	375	1	ATMSG	actin gamma - mous
639	23.8	5	375	1	S11222	actin gamma, cytos
640	23.8	5	375	1	ATDO	actin - slime mold
641	23.8	5	375	1	ATAX	actin - Acanthamoe
642	23.8	5	375	1	ATZM1	actin - maize
643	23.8	5	375	2	S70377	actin - Phaffia rh
644	23.8	5	375	2	S71124	actin beta-1, cyto
645	23.8	5	375	2	S71125	actin beta-2, cyto
646	23.8	5	375	2	S71126	actin beta, cytoso
647	23.8	5	375	2	S42103	actin - Puccinia g
648	23.8	5	375	2	A55001	actin beta - goose
649	23.8	5	375	2	S47897	actin 1 - Pneumocy
650	23.8	5	375	2	A26836	actin - fission ye
651	23.8	5	375	1	JT0385	actin gamma - Emer
652	23.8	5	375	2	S03126	actin - imperfect
653	23.8	5	375	2	A31134	actin, macronuclea
654	23.8	5	375	2	A54728	actin, alpha, cardi
655	23.8	5	375	2	S33386	actin, cytosolic (
656	23.8	5	375	2	T25272	hypothetical prote
657	23.8	5	376	1	ATCHSM	actin gamma, smoot
658	23.8	5	376	2	ATURS	actin Cy1 - sea ur
659	23.8	5	376	2	A43552	actin gamma, cytos
660	23.8	5	376	1	ATFF7	actin 8 - fruit fl
661	23.8	5	376	1	ATFF8	actin 8 - fruit fl
662	23.8	5	376	1	ATFY	actin - slime mold
663	23.8	5	376	2	ATAXE	actin beta - goose
664	23.8	5	376	1	ATRZ3	actin - Entamoeba
665	23.8	5	376	2	ATRZ7	actin 3 - rice
666	23.8	5	376	1	ATSY3	actin 7 - rice
667	23.8	5	376	2	S55892	actin - soybean
668	23.8	5	376	2	S71123	endo-1,4-beta-xyla
669	23.8	5	376	2	S12628	actin alpha-anomal
670	23.8	5	376	2	S07639	actin - malaria pa
671	23.8	5	376	2	A54496	actin - yeast (Can
672	23.8	5	376	2	A54509	actin I - malaria
673	23.8	5	376	2	S49480	actin 3 - Atlantic
674	23.8	5	376	2	JN0832	actin (clone gen3)
675	23.8	5	376	2	S49481	actin 5 - Atlantic
676	23.8	5	376	2	JN0833	actin (clones Ia a
677	23.8	5	376	2	S43509	actin - California
678	23.8	5	376	2	S49479	actin 11 - Atlanti
679	23.8	5	376	2	B23412	actin II - malaria
680	23.8	5	376	2	A25084	actin 12 - slime m
681	23.8	5	376	2	S11453	actin 15 - slime m
682	23.8	5	376	2	JQ0154	actin (clone 403)
683	23.8	5	376	2	S07382	actin - Hydra atte
684	23.8	5	376	2	S09059	actin A2 - silkwor
685	23.8	5	376	2	S12730	actin A1 - silkwor
686	23.8	5	376	2	S04538	actin - California

#			Len		Acc	Description
687	5	23.8	376	2	JS0190	actin, muscle - st
688	5	23.8	376	2	S16709	actin 2 - Caenorha
689	5	23.8	376	2	S27135	actin 4 - Caenorha
690	5	23.8	376	2	C23412	actin 3-sub1 - sil
691	5	23.8	376	2	S11451	actin (clone 211)
692	5	23.8	376	2	S11450	actin (clone 205)
693	5	23.8	376	2	S24408	actin - Achlya bis
694	5	23.8	376	2	A45634	actin - Cryptospor
695	5	23.8	376	2	A29407	actin - Tetrahymen
696	5	23.8	376	2	S07284	actin - Tetrahymen
697	5	23.8	376	2	A48449	Actin-1A - nematod
698	5	23.8	376	2	JC1246	actin - fruit fly
699	5	23.8	376	2	JE0414	actin A - Phytopht
700	5	23.8	376	2	S16710	actin 1 and actin
701	5	23.8	376	2	JS0189	actin, cytosolic -
702	5	23.8	376	2	A25135	actin A3, cytosoli
703	5	23.8	376	2	A32788	actin gamma, smoot
704	5	23.8	376	2	A31375	actin, smooth musc
705	5	23.8	376	2	A40261	actin gamma, enter
706	5	23.8	376	2	S07288	actin 15A - sea ur
707	5	23.8	376	2	S09578	actin - sea urchin
708	5	23.8	376	2	S26435	actin 2 - garden p
709	5	23.8	376	2	S25488	actin 1 - garden p
710	5	23.8	376	2	JC5227	actin 1 - earthwor
711	5	23.8	376	2	JC5228	actin 2 - earthwor
712	5	23.8	376	2	JC5750	actin A4 - silkwor
713	5	23.8	376	2	A44940	actin - pork tapew
714	5	23.8	376	2	T24448	hypothetical prote
715	5	23.8	376	2	T28935	hypothetical prote
716	5	23.8	377	1	ATHU	actin alpha 1, ske
717	5	23.8	377	1	ATRT	actin, skeletal mu
718	5	23.8	377	1	A24904	actin alpha, skele
719	5	23.8	377	1	ATCH	actin alpha, skele
720	5	23.8	377	1	ATHUSM	actin alpha 2, aor
721	5	23.8	377	1	A22224	actin alpha, vascu
722	5	23.8	377	1	ATRBSM	actin alpha, smoot
723	5	23.8	377	1	A25719	actin alpha, aorti
724	5	23.8	377	1	ATHUC	actin, cardiac mus
725	5	23.8	377	1	A23022	actin, cardiac mus
726	5	23.8	377	1	B29686	actin alpha, cardi
727	5	23.8	377	1	ATRZ1	actin 1 - rice
728	5	23.8	377	2	JE0147	actin 1 - sorghum
729	5	23.8	377	2	S65079	actin - Cyanidiosc
730	5	23.8	377	2	S71118	actin alpha-1, ske
731	5	23.8	377	2	S71119	actin alpha-2, ske
732	5	23.8	377	2	S71120	actin alpha, cardi
733	5	23.8	377	2	S14120	actin - Volvox car
734	5	23.8	377	2	S31933	actin - common tob
735	5	23.8	377	2	B24848	actin alpha-3, ske
736	5	23.8	377	2	A24848	actin alpha-1, car
737	5	23.8	377	2	A29686	actin alpha-2, ske
738	5	23.8	377	2	S58316	actin - garden pea
739	5	23.8	377	2	JC4612	actin - Chlamydomo
740	5	23.8	377	2	S20093	actin 101 - potato
741	5	23.8	377	2	S20098	actin 97 - potato
742	5	23.8	377	2	S20094	actin 58 - potato
743	5	23.8	377	2	S20096	actin 75 - potato
744	5	23.8	377	2	S68109	actin 11 - Arabido
745	5	23.8	377	2	S68112	actin 3 [imported]
746	5	23.8	377	2	S68108	actin 12 - Arabido
747	5	23.8	377	2	S68106	actin 4 - Arabidop
748	5	23.8	377	2	S68107	actin 7 - Arabidop
749	5	23.8	377	2	S20095	actin 71 - potato
750	5	23.8	377	2	JC5301	skeletal alpha-act
751	5	23.8	377	2	T51180	actin [imported] -
752	5	23.8	377	2	T51181	actin 1 [imported]
753	5	23.8	377	2	T51182	actin [imported] -
754	5	23.8	377	2	T51183	actin isoform B [i
755	5	23.8	377	2	T51184	actin [imported] -
756	5	23.8	377	2	T51175	actin [imported] -
757	5	23.8	377	2	T51176	actin [imported] -
758	5	23.8	377	2	T51177	actin [imported] -
759	5	23.8	377	2	T51178	actin ACT2 [import

#	Acc		Len			Description
760	ATSY1	1	378	23.8	5	actin 1 - soybean
761	S21907	2	378	23.8	5	actin, muscle - se
762	A43911	2	378	23.8	5	actin 1-alpha, mus
763	JC2142	1	379	23.8	5	alkaline proteinas
764	ATRZ2	1	379	23.8	5	actin 2 - rice
765	S33387	2	379	23.8	5	actin, muscle - se
766	T39400	2	379	23.8	5	actin like protein
767	T26246	2	379	23.8	5	hypothetical prote
768	H96548	2	379	23.8	5	unknown protein [i
769	F81449	2	380	23.8	5	8-amino-7-oxononan
770	B64533	2	380	23.8	5	cystathionine gamm
771	D71973	2	380	23.8	5	probable cystathio
772	S07002	2	380	23.8	5	actin 1 - carrot
773	S07003	2	380	23.8	5	actin 2 - carrot
774	G96804	2	381	23.8	5	hypothetical prote
775	F96009	2	382	23.8	5	hypothetical membr
776	G83808	2	382	23.8	5	hypothetical prote
777	E96499	2	383	23.8	5	hypothetical prote
778	S76812	2	383	23.8	5	hypothetical prote
779	S66292	2	385	23.8	5	actin-crosslinking
780	H71960	2	385	23.8	5	hypothetical prote
781	T06788	2	386	23.8	5	actin - garden pea
782	S76131	2	386	23.8	5	nitrogenase cofact
783	A45117	2	387	23.8	5	aspartic proteinas
784	A97709	2	387	23.8	5	hypothetical prote
785	C32905	2	388	23.8	5	desmoplakin, desmo
786	S41748	2	388	23.8	5	heat shock protein
787	C82130	2	389	23.8	5	conserved hypothet
788	C90215	2	389	23.8	5	molybdopterin bios
789	H86656	2	389	23.8	5	hypothetical prote
790	D64548	1	390	23.8	5	integral membrane
791	RRNZMS	2	390	23.8	5	polymerase-associa
792	C95160	2	390	23.8	5	conserved hypothet
793	B84584	2	390	23.8	5	probable RING zinc
794	G89877	2	391	23.8	5	conserved hypothet
795	RRNZYA	2	391	23.8	5	polymerase-associa
796	RRNZED	2	391	23.8	5	polymerase-associa
797	AB0443	2	391	23.8	5	hypothetical prote
798	E71633	2	392	23.8	5	acetyl-COA acetylt
799	A84125	2	392	23.8	5	acetyl-COA acetylt
800	T38450	2	392	23.8	5	probable GTP-bindi
801	G1MSM	1	393	23.8	5	Ig gamma-1 chain C
802	T31121	2	394	23.8	5	hypothetical prote
803	C82191	2	395	23.8	5	phospho-2-dehydro-
804	T05680	2	395	23.8	5	hypothetical prote
805	S70912	2	395	23.8	5	CMP-N-acetylneuram
806	G96917	2	395	23.8	5	related to ABC tra
807	E64143	1	396	23.8	5	chloramphenicol re
808	S39674	1	396	23.8	5	ywbD protein - Bac
809	T47207	2	396	23.8	5	aspartic proteinas
810	S54641	2	396	23.8	5	probable membrane
811	T01201	2	396	23.8	5	hypothetical prote
812	AB2309	2	396	23.8	5	hypothetical prote
813	S61066	2	397	23.8	5	6-phosphofructo-2-
814	F90182	2	397	23.8	5	hypothetical prote
815	E81223	2	397	23.8	5	conserved hypothet
816	B81994	2	397	23.8	5	probable transmemb
817	E86295	2	397	23.8	5	hypothetical prote
818	T20901	2	398	23.8	5	hypothetical prote
819	C91063	2	398	23.8	5	hypothetical prote
820	A57261	2	398	23.8	5	probable transposa
821	D95849	2	398	23.8	5	probable ISRm5 tra
822	G2MSAM	1	399	23.8	5	Ig gamma-2a chain
823	E86289	2	399	23.8	5	Tf6N11.7 protein -
824	JC4265	2	400	23.8	5	plasminogen activa
825	T18060	2	400	23.8	5	hypothetical prote
826	T24381	2	401	23.8	5	hypothetical prote
827	ITHUP1	2	402	23.8	5	plasminogen activa
828	S06745	1	402	23.8	5	plasminogen activa
829	D86575	2	402	23.8	5	phosphoglycerate k
830	C72049	2	402	23.8	5	phosphoglycerate k
831	G83289	2	402	23.8	5	conserved hypothet
832	B89819	2	404	23.8	5	pyrimidine nucleos

#			Len		Accession	N	Description
833	5	23.8	405	23.8	G2MSBM	1	Ig gamma-2b chain
834	5	23.8	407	23.8	KHRTD	1	cathepsin D (EC 3.
835	5	23.8	408	23.8	VCMVSR	1	env polyprotein -
836	5	23.8	408	23.8	B81417	2	hypothetical prote
837	5	23.8	409	23.8	B75010	2	hypothetical prote
838	5	23.8	410	23.8	KHMSD	1	cathepsin D (EC 3.
839	5	23.8	410	23.8	AD1080	2	arginine deiminase
840	5	23.8	410	23.8	T21960	2	hypothetical prote
841	5	23.8	411	23.8	H95328	2	TRm5 transposase [
842	5	23.8	411	23.8	I55604	1	platelet glycoprot
843	5	23.8	412	23.8	H64239	2	UV protection prot
844	5	23.8	412	23.8	KHHUD	1	cathepsin D (EC 3.
845	5	23.8	413	23.8	T05285	2	farnesyl-diphospha
846	5	23.8	412	23.8	T32577	2	hypothetical prote
847	5	23.8	413	23.8	T44924	2	farnesyl-diphospha
848	5	23.8	413	23.8	E85408	2	probable squalene
849	5	23.8	413	23.8	AC0834	2	probable membrane
850	5	23.8	413	23.8	AH0780	2	probable oxidoredu
851	5	23.8	415	23.8	B55563	2	AML2a protein - hu
852	5	23.8	415	23.8	S60078	1	Runt domain contai
853	5	23.8	415	23.8	F96499	2	hypothetical prote
854	5	23.8	415	23.8	T32467	2	hypothetical prote
855	5	23.8	416	23.8	T44842	2	probable UDPglucos
856	5	23.8	416	23.8	E81352	2	probable zinc prot
857	5	23.8	416	23.8	D75531	2	folyl-polyglutamat
858	5	23.8	416	23.8	B55649	2	TNFR-associated pr
859	5	23.8	416	23.8	S75793	2	poly(A) polymerase
860	5	23.8	417	23.8	F86731	2	exodeoxyribonuclea
861	5	23.8	417	23.8	T33207	2	hypothetical prote
862	5	23.8	417	23.8	AD0907	2	cytosine permease
863	5	23.8	418	23.8	JC4069	2	rfbX protein - Shi
864	5	23.8	420	23.8	F84142	2	glutamate dehydrog
865	5	23.8	421	23.8	E72573	2	probable 3-hydroxy
866	5	23.8	421	23.8	D95975	2	hypothetical outer
867	5	23.8	422	23.8	T09742	2	drought-induced pr
868	5	23.8	423	23.8	F75635	2	hypothetical prote
869	5	23.8	424	23.8	SYECYT	1	tyrosine-tRNA lig
870	5	23.8	424	23.8	B90922	2	tyrosine tRNA synt
871	5	23.8	424	23.8	AE0693	2	tyrosyl-tRNA synth
872	5	23.8	426	23.8	T45767	2	hypothetical prote
873	5	23.8	427	23.8	T45915	1	actin (ACT3) - Ara
874	5	23.8	427	23.8	B86408	2	probable dnaJ prot
875	5	23.8	427	23.8	T39430	2	mitochondrial impo
876	5	23.8	428	23.8	AD3593	2	dihydroorotase (EC
877	5	23.8	428	23.8	F85770	2	tyrosine tRNA synt
878	5	23.8	428	23.8	AD0085	2	probable type II s
879	5	23.8	430	23.8	S32570	1	malC protein - Str
880	5	23.8	431	23.8	S52583	2	crtX protein - Erw
881	5	23.8	431	23.8	E84452	2	hypothetical prote
882	5	23.8	432	23.8	JN0317	1	protein-tyrosine-p
883	5	23.8	432	23.8	A34845	1	protein-tyrosine-p
884	5	23.8	432	23.8	A25483	2	env polyprotein, r
885	5	23.8	432	23.8	H64152	2	hypothetical prote
886	5	23.8	433	23.8	C64605	1	UDP-N-acetylglucos
887	5	23.8	433	23.8	A71909	2	udp-n-acetylglucos
888	5	23.8	435	23.8	TPHUN1	1	protein-tyrosine-p
889	5	23.8	435	23.8	G95246	2	maltodextrin ABC t
890	5	23.8	435	23.8	D98111	2	hypothetical prote
891	5	23.8	436	23.8	E82848	2	alpha-ketoglutarat
892	5	23.8	436	23.8	B84324	2	glutamyl-tRNA redu
893	5	23.8	436	23.8	B96828	2	probable aspartyl
894	5	23.8	436	23.8	S77639	2	exopolysaccharide
895	5	23.8	438	23.8	T03513	2	probable secretion
896	5	23.8	439	23.8	E75118	2	probable threonine
897	5	23.8	440	23.8	T31458	2	protochlorophyllid
898	5	23.8	440	23.8	H71013	1	probable threonine
899	5	23.8	440	23.8	D95029	2	PTS system, IIC co
900	5	23.8	440	23.8	G97900	2	hypothetical prote
901	5	23.8	441	23.8	T01613	2	hypothetical prote
902	5	23.8	441	23.8	F86185	2	hypothetical prote
903	5	23.8	442	23.8	H91097	2	hypothetical prote
904	5	23.8	442	23.8	H72209	2	hypothetical prote
905	5	23.8	442	23.8	S72568	2	hypothetical prote
906	5	23.8	442	23.8	D87351	2	conserved hypothet
907	5	23.8	443	23.8	S76611	1	hypothetical prote
908	5	23.8	443	23.8	AE0826	2	probable cadaverin
909	5	23.8	443	23.8	H70430	2	K+ transport prote
910	5	23.8	443	23.8	S65963	2	flavastacin (EC 3.
911	5	23.8	444	23.8	PC4436	2	monoclonal antibod
912	5	23.8	444	23.8	A41842	2	lysine/cadaverine
913	5	23.8	444	23.8	B91268	2	transport protein
914	5	23.8	444	23.8	H86108	2	transport of lysin
915	5	23.8	444	23.8	AC3162	2	conserved hypothet
916	5	23.8	445	23.8	G95204	2	conserved hypothet
917	5	23.8	446	23.8	S40295	2	Ig gamma-2a chain
918	5	23.8	446	23.8	D89811	2	hypothetical prote
919	5	23.8	447	23.8	B75212	2	amidophosphoribosy
920	5	23.8	447	23.8	AE1026	2	hypothetical prote
921	5	23.8	447	23.8	H84797	2	hypothetical prote
922	5	23.8	449	23.8	A71248	2	amidophosphoribosy
923	5	23.8	449	23.8	F64536	2	3-deoxy-D-arabino-
924	5	23.8	449	23.8	D71971	2	phospho-2-dehydro-
925	5	23.8	449	23.8	D82215	2	sensor histidine k
926	5	23.8	449	23.8	F97379	2	phoT protein U5922
927	5	23.8	449	23.8	AE2597	2	hypothetical prote
928	5	23.8	450	23.8	B81730	2	conserved hypothet
929	5	23.8	451	23.8	F69825	2	sodium-dependent t
930	5	23.8	451	23.8	T41911	2	hypothetical prote
931	5	23.8	451	23.8	T49095	2	beta-ketoacyl-CoA
932	5	23.8	451	23.8	T31790	2	hypothetical prote
933	5	23.8	452	23.8	F91292	2	hypothetical prote
934	5	23.8	452	23.8	A86134	2	hypothetical prote
935	5	23.8	453	23.8	F90327	2	cytosine permease
936	5	23.8	458	23.8	B81409	2	probable transmemb
937	5	23.8	459	23.8	T45576	2	anthranilate N-hyd
938	5	23.8	461	23.8	G83602	2	hypothetical prote
939	5	23.8	461	23.8	A46394	2	suppressor protein
940	5	23.8	463	23.8	P2WL13	1	L2 protein - human
941	5	23.8	463	23.8	E83156	2	conserved hypothet
942	5	23.8	463	23.8	T36810	2	probable integral
943	5	23.8	464	23.8	MNVUWC	1	nonstructural prot
944	5	23.8	464	23.8	AG1146	2	beta-glucosidase h
945	5	23.8	464	23.8	AG1505	2	beta-glucosidase h
946	5	23.8	464	23.8	G65071	2	hypothetical prote
947	5	23.8	464	23.8	D85943	2	probable proteogly
948	5	23.8	465	23.8	T21505	2	hypothetical prote
949	5	23.8	465	23.8	D69785	2	beta-glucosidase h
950	5	23.8	466	23.8	F84139	2	beta-glucosidase B
951	5	23.8	466	23.8	T44650	2	capsular polysacch
952	5	23.8	466	23.8	T08914	2	hypothetical prote
953	5	23.8	466	23.8	F95307	2	conserved hypothet
954	5	23.8	467	23.8	PWBSBF	1	H+-transporting tw
955	5	23.8	467	23.8	MNVUW1	1	nonstructural prot
956	5	23.8	468	23.8	T03164	2	probable major env
957	5	23.8	468	23.8	S61964	2	probable membrane
958	5	23.8	469	23.8	AC3650	2	6-phosphogluconate
959	5	23.8	469	23.8	D86144	2	protein probable U
960	5	23.8	469	23.8	S37483	2	Ig gamma-2a chain
961	5	23.8	472	23.8	A53048	2	translation initia
962	5	23.8	474	23.8	G2MS11	1	Ig gamma-2b chain
963	5	23.8	474	23.8	T37967	2	4-aminobutyrate tr
964	5	23.8	474	23.8	A35732	2	protective protein
965	5	23.8	474	23.8	T20108	2	hypothetical prote
966	5	23.8	475	23.8	S46941	2	translation initia
967	5	23.8	475	23.8	S01321	2	Ig gamma-2b chain
968	5	23.8	478	23.8	H71224	2	hypothetical prote
969	5	23.8	478	23.8	T31955	2	hypothetical prote
970	5	23.8	479	23.8	S61768	2	origin recognition
971	5	23.8	480	23.8	A31589	2	carboxypeptidase C
972	5	23.8	480	23.8	H82506	2	hypothetical prote
973	5	23.8	480	23.8	F82796	2	hypothetical prote
974	5	23.8	481	23.8	B84700	2	probable flavonol
975	5	23.8	481	23.8	E83062	2	deoxyribodipyrimid
976	5	23.8	481	23.8	G75192	2	probable ss-DNA-sp
977	5	23.8	481	23.8	T27435	1	hypothetical prote
978	5	23.8	482	23.8	A26934	1	nitrogen assimilat

```
979   5  23.8  482  2  F86902  dextransucrase (EC
980   5  23.8  483  2  G86902  dextransucrase (EC
981   5  23.8  484  2  H69298  hypothetical prote
982   5  23.8  485  2  S74708  hypothetical prote
983   5  23.8  485  2  S34800  sulfate transport
984   5  23.8  487  2  AE0310  NADH dehydrogenase
985   5  23.8  487  2  S65133  butyrophilin - mou
986   5  23.8  488  2  T31296  glycolate oxidase
987   5  23.8  490  2  C71541  probable na-depend
988   5  23.8  490  2  E69066  conserved hypothet
989   5  23.8  490  2  H83286  probable chemotaxi
990   5  23.8  490  2  B97221  L-fucose isomerase
991   5  23.8  491  2  I40455  penicillin binding
992   5  23.8  492  2  S46298  catalase (EC 1.11.
993   5  23.8  493  2  S49147  ERF-2 protein - hu
994   5  23.8  495  2  T25357  hypothetical prote
995   5  23.8  495  2  D64578  conserved hypothet
996   5  23.8  498  1  JQ0197  4-aminobutyrate tr
997   5  23.8  498  2  T39077  hypothetical prote
998   5  23.8  498  2  F83329  probable outer mem
999   5  23.8  498  2  S28059  serum response fac
1000  5  23.8  499  2  S70113  hypothetical prote
```

ALIGNMENTS

RESULT 1
S46737
hypothetical protein YHR038w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein H8179.10
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C;Accession: S46737
R;Du, Z.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 8179.
A;Reference number: S46732
A;Accession: S46737
A;Molecule type: DNA
A;Residues: 1-230 <DUZ>
A;Cross-references: EMBL:U00062; NID:g488162; PID:g488171; GSPDB:GN00008; MIPS:YHR038w
C;Genetics:
A;Gene: MIPS:YHR038w
A;Map position: 8R

```
Query Match       33.3%;  Score 7;  DB 2;  Length 230;
Best Local Similarity  100.0%;  Pred. No. 3.7;
Matches   7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   9 RLFNRSF 15
        |||||||
Db  17 RLFNRSF 23
```

RESULT 2
S70367
stem cell factor short form precursor - quail
C;Species: Coturnix coturnix (quail)
C;Date: 06-Dec-1996 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
C;Accession: S70367
R;Petitte, J.N.; Kulik, M.J.
Biochim. Biophys. Acta 1307, 149-151, 1996
A;Title: Cloning and characterization of cDNAs encoding two forms of avian stem cell fac
A;Reference number: S70366; MUID:96283808
A;Accession: S70367
A;Molecule type: mRNA
A;Residues: 1-253 <PET>
A;Cross-references: EMBL:U43079; NID:g1150877; PIDN:AAC59934.1; PID:g1150878
C;Superfamily: mouse mast cell growth factor
C;Keywords: growth factor; transmembrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-253/Product: stem cell factor short form #status predicted <MAT>
F;192-216/Domain: transmembrane #status predicted <TMM>

```
Query Match       33.3%;  Score 7;  DB 2;  Length 253;
Best Local Similarity  100.0%;  Pred. No. 4.1;
Matches   7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   7 FFRLFNR 13
        |||||||
Db 144 FFRLFNR 150
```

RESULT 3
S70366
stem cell factor long form precursor - quail
C;Species: Coturnix coturnix (quail)
C;Date: 06-Dec-1996 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
C;Accession: S70366
R;Petitte, J.N.; Kulik, M.J.
Biochim. Biophys. Acta 1307, 149-151, 1996
A;Title: Cloning and characterization of cDNAs encoding two forms of avian stem cell
A;Reference number: S70366; MUID:96283808
A;Accession: S70366
A;Molecule type: mRNA
A;Residues: 1-287 <PET>
A;Cross-references: EMBL:U43078; NID:g1150875; PIDN:AAC59933.1; PID:g1150876
C;Superfamily: mouse mast cell growth factor
C;Keywords: growth factor; transmembrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-287/Product: stem cell factor long form #status predicted <MAT>
F;226-250/Domain: transmembrane #status predicted <TMM>

```
Query Match       33.3%;  Score 7;  DB 2;  Length 287;
Best Local Similarity  100.0%;  Pred. No. 4.6;
Matches   7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   7 FFRLFNR 13
        |||||||
Db 144 FFRLFNR 150
```

RESULT 4
AD2284
hypothetical protein alr3827 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AD2284
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2284
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75526.1; PID:g17132961; GSPDB:GN0179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3827
C;Superfamily: Synechocystis hypothetical protein sll0827

```
Query Match       33.3%;  Score 7;  DB 2;  Length 472;
Best Local Similarity  100.0%;  Pred. No. 7.2;
Matches   7;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  14 SFTQALG 20
        |||||||
Db 408 SFTQALG 414
```

RESULT 5
AI0479

probable membrane protein YPO3943 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AI0479
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AI0479
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-565 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93405.1; PID:g15981851; GSPDB:GN00175
C;Genetics:
A;Gene: YPO3943
C;Superfamily: hypothetical protein b2380

Query Match   33.3%; Score 7; DB 2; Length 565;
Best Local Similarity   100.0%; Pred. No. 8.4;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3 SLSTFFR 9
          |||||||
Db   410 SLSTFFR 416

RESULT 6
A97416
hypothetical protein AGR_C_830 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: A97416
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: A97416
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-43 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86282.1; PID:g15155394; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_830
A;Map position: circular chromosome

Query Match   28.6%; Score 6; DB 2; Length 43;
Best Local Similarity   100.0%; Pred. No. 9.6;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2 GSLSTF 7
          ||||||
Db   31 GSLSTF 36

RESULT 7
T18094
hypothetical protein A592R - Chlorella virus PBCV-1
C;Species: Chlorella virus PBCV-1
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18094
R;Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18806
A;Accession: T18094
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-69 <GRA>
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96934.1
A;Experimental source: specific host Chlorella strain NC64A
C;Genetics:

A;Note: A592R

Query Match   28.6%; Score 6; DB 2; Length 69;
Best Local Similarity   100.0%; Pred. No. 15;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3 SLSTFFF 8
          |||||||
Db   27 SLSTFFF 32

RESULT 8
B75387
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: B75387
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-122 <WHI>
A;Cross-references: GB:AE001995; GB:AE005513; NID:g6459271; PIDN:AAF11085.1; PID:g645
C;Genetics:
A;Gene: DR1518
A;Map position: 1
C;Superfamily: Deinococcus radiodurans hypothetical protein DR1518

Query Match   28.6%; Score 6; DB 2; Length 122;
Best Local Similarity   100.0%; Pred. No. 25;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
          ||||||
Db   66 SGSLST 71

RESULT 9
AE1330
hypothetical protein lmo2045 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1330
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1330
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-128 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00123.1; PID:g16411515; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2045

Query Match   28.6%; Score 6; DB 2; Length 128;
Best Local Similarity   100.0%; Pred. No. 26;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   10 LFNRSF 15
          ||||||
Db   60 LFNRSF 65

RESULT 10
T17655
hypothetical protein a164L - Chlorella virus PBCV-1
C;Species: Chlorella virus PBCV-1
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17655
R;Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18806
A;Accession: T17655
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-134 <GRA>
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96532.1
A;Experimental source: specific host Chlorella strain NC64A
C;Genetics:
A;Note: a164L

Query Match   28.6%; Score 6; DB 2; Length 134;
Best Local Similarity   100.0%; Pred. No. 27;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   11 FNRSFT 16
            ||||||
Db   111 FNRSFT 116

RESULT 11
B70370
hypothetical protein aq_805 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: B70370
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: B70370
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <AQF>
A;Cross-references: GB:AE000708; NID:g2983356; PIDN:AAC06954.1; PID:g2983367; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_805
C;Superfamily: Aquifex aeolicus hypothetical protein aq_805

Query Match   28.6%; Score 6; DB 2; Length 155;
Best Local Similarity   100.0%; Pred. No. 31;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   10 LFNRSF 15
            ||||||
Db   54 LFNRSF 59

RESULT 12
S78572
EST3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YIL009c-a
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 04-Mar-2000
C;Accession: S78572
R;Rowley, N.
submitted to the EMBL Data Library, August 1994
A;Reference number: S48442
A;Accession: S78572
A;Molecule type: DNA
A;Residues: 1-181 <ROW>
A;Cross-references: EMBL:Z38113; GSPDB:GN00009; MIPS:YIL009c-a

C;Genetics:
A;Gene: EST3; MIPS:YIL009c-a
A;Cross-references: MIPS:YIL009c-a
A;Map position: 9L
A;Introns: 92/3
C;Function:
A;Description: required for telomerase function
C;Superfamily: Saccharomyces cerevisiae EST3 protein

Query Match   28.6%; Score 6; DB 2; Length 181;
Best Local Similarity   100.0%; Pred. No. 35;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2 GSLSTF 7
          ||||||
Db   169 GSLSTF 174

RESULT 13
A56956
GTP-binding protein-associated protein (clone B) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 05-Nov-1999
C;Accession: A56956
R;Janoueix-Lerosey, I.; Jollivet, F.; Camonis, J.; Marche, P.N.; Goud, B.
J. Biol. Chem. 270, 14801-14808, 1995
A;Title: Two-hybrid system screen with the small GTP-binding protein Rab6. Identifica
A;Reference number: A56956; MUID:95301579
A;Accession: A56956
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-186 <JAN>
A;Cross-references: GB:L40894; NID:g725273; PIDN:AAA78787.1; PID:g725274

Query Match   28.6%; Score 6; DB 2; Length 186;
Best Local Similarity   100.0%; Pred. No. 36;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
          ||||||
Db   130 SGSLST 135

RESULT 14
A44363
voltage-gated dihydropyridine-sensitive calcium channel subtype homolog (internal rep
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 11-Jan-2000
C;Accession: A44363
R;Sun, W.; McPherson, J.D.; Hoang, D.Q.; Wasmuth, J.J.; Evans, G.A.; Montal, M.
Genomics 14, 1092-1094, 1992
A;Title: Mapping of a human brain voltage-gated calcium channel to human chromosome 1
A;Reference number: A44363; MUID:93122776
A;Accession: A44363
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-197 <SUN>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIN:122411, NCBIP:122412)
C;Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match   28.6%; Score 6; DB 2; Length 197;
Best Local Similarity   100.0%; Pred. No. 38;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   6 TFFRLF 11
          ||||||
Db   150 TFFRLF 155

RESULT 15
H81731
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase TC0180 [imported]

C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C;Accession: H81731
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255
A;Accession: H81731
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-202 <TET>
A;Cross-references: GB:AE002285; GB:AE002160; NID:g7190212; PIDN:AAF39054.1; PID:g7190212|
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0180
C;Superfamily: CDPdiacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase

Query Match   28.6%;   Score 6;   DB 2;   Length 202;
Best Local Similarity   100.0%;   Pred. No. 39;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   6 TFFRLF 11
          ||||||
Db   10 TFFRLF 15

Search completed: November 5, 2002, 11:07:10
Job time : 21.1493 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2002 Compugen Ltd.
GenCore version 5.1.3

Run on: November 5, 2002, 10:58:02 ; Search time 5.64179 Seconds
                                           (without alignments)
                                           144.123 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 21
Sequence: 1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table: OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size :   0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   SwissProt_40:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	33.3	230	1	FIL1_YEAST	P38771 saccharomyc
2	7	33.3	500	1	ABF1_TRIRE	Q92455 trichoderma
3	7	33.3	500	1	XYL1_TRIKO	P48792 trichoderma
4	6	28.6	155	1	Y805_AQUAE	O66989 aquifex aeo
5	6	28.6	169	1	CCAC_CAVPO	O35505 cavia porce
6	6	28.6	177	1	CCAC_CHICK	O73707 gallus gall
7	6	28.6	214	1	Y540_AQUAE	O68819 aquifex aeo
8	6	28.6	281	1	CCAD_MOUSE	Q99246 mus musculu
9	6	28.6	281	1	CCAS_CHICK	O42398 gallus gall
10	6	28.6	287	1	SCF_CHICK	Q09108 gallus gall
11	6	28.6	296	1	CEBB_MOUSE	P28033 mus musculu
12	6	28.6	297	1	CEBB_RAT	P21272 rattus norv
13	6	28.6	329	1	ISPB_HAEIN	P44916 haemophilus
14	6	28.6	345	1	CEBB_HUMAN	P17676 homo sapien
15	6	28.6	348	1	CEBB_BOVIN	O02755 bos taurus
16	6	28.6	354	1	FXR1_HUMAN	Q12946 homo sapien
17	6	28.6	373	1	ACT6_DIPDE	P53459 diphyllobot
18	6	28.6	405	1	WCAC_ECOLI	P71237 escherichia
19	6	28.6	499	1	ABFB_ASPNG	P42255 aspergillus
20	6	28.6	561	1	YBHU_ECOLI	P33357 escherichia
21	6	28.6	634	1	SL56_RAT	O70247 rattus norv
22	6	28.6	635	1	SL56_HUMAN	Q9y289 homo sapien
23	6	28.6	636	1	SL56_RABIT	Q9xt77 oryctolagus
24	6	28.6	637	1	MUTA_PROFR	P11652 propionibac
25	6	28.6	699	1	YQQA_CAEEL	Q09299 caenorhabdi
26	6	28.6	759	1	PMT6_YEAST	P42934 saccharomyc
27	6	28.6	841	1	YAGX_ECOLI	P77802 escherichia
28	6	28.6	970	1	Y087_BUCAI	P57189 buchnera ap
29	6	28.6	1075	1	PST2_SCHPO	O13919 schizosacch
30	6	28.6	1126	1	RPOB_SULAC	P11513 sulfolobus
31	6	28.6	1610	1	CCAD_MESAU	Q99244 mesocricetu
32	6	28.6	1787	1	UVRA_CHLMU	Q9pk60 chlamydia m
33	6	28.6	1807	1	ITB4_RAT	Q64632 rattus norv
34	6	28.6	1852	1	CCAS_CYPCA	P22316 cyprinus ca
35	6	28.6	1966	1	CCAF_HUMAN	060840 homo sapien
36	6	28.6	2139	1	CCAC_MOUSE	Q01815 mus musculu
37	6	28.6	2161	1	CCAD_HUMAN	Q01668 homo sapien
38	6	28.6	2169	1	CCAC_RAT	P22002 rattus norv
39	6	28.6	2171	1	CCAC_RABIT	P15381 oryctolagus
40	6	28.6	2190	1	CCAD_CHICK	073700 gallus gall
41	6	28.6	2203	1	CCAD_RAT	P27732 rattus norv
42	6	28.6	2221	1	CCAC_HUMAN	Q13936 homo sapien
43	5	23.8	33	1	ACT_DICVI	Q24733 dictyocaulu
44	5	23.8	67	1	YF82_ARCFU	028690 archaeoglob
45	5	23.8	87	1	CYSD_CHLLT	P20958 chlorobium
46	5	23.8	88	1	GP42_BPSP1	048398 bacteriopha
47	5	23.8	94	1	VA15_VACCC	P20992 vaccinia vi
48	5	23.8	94	1	VA15_VARV	P33840 variola vir
49	5	23.8	97	1	YC40_METJA	Q58637 methanococc
50	5	23.8	102	1	YAE5_YEAST	P39725 saccharomyc
51	5	23.8	104	1	YGY9_YEAST	P53069 saccharomyc
52	5	23.8	109	1	FEL2_FELCA	P30440 felis silve
53	5	23.8	110	1	VL1_FPVL	P06456 avian papil
54	5	23.8	113	1	GVK1_HALN1	P24375 halobacteri
55	5	23.8	113	1	Y959_CAMJE	Q9pnx6 campylobact
56	5	23.8	117	1	GHRL_CANFA	Q9bef8 canis famil
57	5	23.8	117	1	GHRL_HUMAN	Q9obu3 homo sapien
58	5	23.8	117	1	GVPK_HALME	Q02236 halobacteri
59	5	23.8	119	1	GVK2_HALN1	P33963 halobacteri
60	5	23.8	120	1	RBFA_BUCAI	P57457 buchnera ap
61	5	23.8	120	1	YE35_PYRHO	059124 pyrococcus
62	5	23.8	125	1	ACTS_PLEWA	P10994 pleurodeles
63	5	23.8	128	1	IF2G_RABIT	P33887 oryctolagus
64	5	23.8	130	1	FABP_BLOTA	Q17284 blomia trop
65	5	23.8	130	1	GLOV_HYACE	P81048 hyalophora
66	5	23.8	131	1	FABP_LEPDS	Q9uspl lepidoglyph
67	5	23.8	139	1	A85A_MYCMR	Q9xh57 mycobacteri
68	5	23.8	140	1	Y337_MYCPN	P75297 mycoplasma
69	5	23.8	140	1	YC49_ARCFU	029019 archaeoglob
70	5	23.8	141	1	YEF5_YEAST	P32616 saccharomyc
71	5	23.8	148	1	YEAL_ECOLI	P76240 escherichia
72	5	23.8	151	1	SP2C_STRPU	P04111 strongyloce
73	5	23.8	154	1	ACT4_LYTPI	Q25380 lytechinus
74	5	23.8	155	1	Y359_RICPR	Q9zdg9 rickettsia
75	5	23.8	161	1	ACT_PINCO	P24902 pinus conto
76	5	23.8	161	1	Y311_RICPR	Q9zd16 rickettsia
77	5	23.8	167	1	LSPA_CHLMU	Q9pjy8 chlamydia m
78	5	23.8	169	1	CALB_HUMAN	P06705 homo sapien
79	5	23.8	169	1	CALB_MOUSE	063810 mus musculu
80	5	23.8	172	1	ACT3_LYTPI	Q25379 lytechinus
81	5	23.8	172	1	ACTM_LYTPI	Q25381 lytechinus
82	5	23.8	173	1	YFHC_HAEIN	P44931 haemophilus
83	5	23.8	175	1	CALC_RAT	P28470 rattus norv
84	5	23.8	177	1	ATPD_BUCAP	051875 buchnera ap
85	5	23.8	178	1	CALC_MOUSE	063811 mus musculu
86	5	23.8	180	1	ARF_PLAFO	Q25761 plasmodium
87	5	23.8	181	1	ARF_PLAFA	094650 plasmodium
88	5	23.8	188	1	Y65L_HORVU	048609 hordeum vul
89	5	23.8	193	1	COME_METJA	P58416 methanococc
90	5	23.8	203	1	YBF9_YEAST	P34224 saccharomyc
91	5	23.8	204	1	KTHY_CHLMU	Q9pkk5 chlamydia m
92	5	23.8	204	1	CAHB_MOUSE	070354 mus musculu
93	5	23.8	205	1	FADD_MOUSE	Q61160 mus musculu
94	5	23.8	207	1	3MGH_LISIN	Q92089 listeria in
95	5	23.8	207	1	3MGH_LISMO	P58621 listeria mo
96	5	23.8	214	1	ACH2_LONAC	Q23605 lonomia ach
97	5	23.8	214	1	YQ37_MYCLE	049642 mycobacteri
98	5	23.8	218	1	YQ37_MYCTU	P71936 mycobacteri
99	5	23.8	222	1	RISB_SPIOL	Q9xh32 spinacia ol
100	5	23.8	224	1	Y08B_MYCTU	Q10885 mycobacteri
101	5	23.8	233	1	DEOD_TREPA	083716 treponema p
102	5	23.8	233	1	YSR3_CAEEL	Q09951 caenorhabdi
103	5	23.8	236	1	ABME_HUMAN	P41238 homo sapien
104	5	23.8	239	1	VIB1_AGRT9	P05350 agrobacteri
105	5	23.8	243	1	GUB_BACLI	P27051 bacillus li
106	5	23.8	247	1	RL7_ARATH	Q42208 arabidopsis

Idx	Score	Len	Name	Num	N	Accession	Species
107	23.8	5	NUKC_MAIZE	248	1	P06670	zea mays (m
108	23.8	5	YKC6_CAEEL	252	1	P41997	caenorhabdi
109	23.8	5	YP96_CAEEL	256	1	Q09476	caenorhabdi
110	23.8	5	Y453_AQUAE	257	1	O66760	aquifex aeo
111	23.8	5	ASH2_MOUSE	262	1	O35885	mus musculu
112	23.8	5	RL7A_FUGRU	265	1	O57592	fugu rubrip
113	23.8	5	RL7A_HUMAN	265	1	P11518	homo sapien
114	23.8	5	RL7A_MOUSE	265	1	P12970	mus musculu
115	23.8	5	YM40_MYCTU	265	1	Q10522	mycobacteri
116	23.8	5	SP2G_CLOAB	266	1	Q45832	clostridium
117	23.8	5	AROE_NEICI	269	1	P95337	neisseria c
118	23.8	5	IL1B_HUMAN	269	1	P01584	homo sapien
119	23.8	5	HPCB_ECOLI	275	1	Q05353	escherichia
120	23.8	5	ACT_CALFI	278	1	Q92192	calanus fin
121	23.8	5	NAE1_THEMA	281	1	Q9wzb3	thermotoga
122	23.8	5	YNFH_ECOLI	284	1	P76173	escherichia
123	23.8	5	Y539_AQUAE	285	1	O66818	aquifex aeo
124	23.8	5	CFA1_MYCTU	287	1	Q11195	mycobacteri
125	23.8	5	EFTS_TREPA	290	1	O83614	treponema p
126	23.8	5	IBP3_MOUSE	291	1	P47878	mus musculu
127	23.8	5	SUD1_ARCFU	291	1	O28733	archaeoglob
128	23.8	5	END4_BACHD	298	1	Q9kd33	bacillus ha
129	23.8	5	PHAG_PSEAE	300	1	Q51553	pseudomonas
130	23.8	5	YHCI_HAEIN	300	1	P44541	haemophilus
131	23.8	5	SC14_KLULA	301	1	P24859	kluyveromyc
132	23.8	5	YCAN_ECOLI	302	1	P75836	escherichia
133	23.8	5	PYRB_BACHD	305	1	Q9k9v6	bacillus ha
134	23.8	5	YEDI_ECOLI	305	1	P46125	escherichia
135	23.8	5	CIW8_MOUSE	307	1	Q9z2t1	mus musculu
136	23.8	5	ACT3_ECHGR	309	1	Q03342	echinococcu
137	23.8	5	NODI_RHILV	311	1	P08720	rhizobium l
138	23.8	5	RT02_ACACA	312	1	P46753	acanthamoeb
139	23.8	5	NOD1_RHILP	314	1	P23718	rhizobium l
140	23.8	5	Y678_METJA	320	1	Q58091	methanococc
141	23.8	5	ACT_PROCL	322	1	P45521	procambarus
142	23.8	5	CITG_HAEDU	323	1	O30827	haemophilus
143	23.8	5	GC1_MOUSE	324	1	P01868	mus musculu
144	23.8	5	ACT3_ARTSX	327	1	P18602	artemia sp.
145	23.8	5	CAHB_HUMAN	328	1	O75493	homo sapien
146	23.8	5	CAHB_SHEEP	328	1	Q95203	ovis aries
147	23.8	5	YCDU_ECOLI	328	1	P75910	escherichia
148	23.8	5	GCAA_MOUSE	330	1	P01863	mus musculu
149	23.8	5	ACT_COSCS	331	1	P30161	costaria co
150	23.8	5	ACT1_SOLTU	332	1	P93587	solanum tub
151	23.8	5	TAL2_YEAST	333	1	P53228	saccharomyc
152	23.8	5	YZ37_SYNY3	333	1	P54480	synechocyst
153	23.8	5	GCAB_MOUSE	335	1	P01864	mus musculu
154	23.8	5	GVPN_ANAFL	335	1	P55150	anabaena fl
155	23.8	5	ACT1_LYCES	336	1	Q96482	lycopersico
156	23.8	5	ACT2_LYCES	336	1	Q96483	lycopersico
157	23.8	5	ACT2_SOLTU	336	1	P93586	solanum tub
158	23.8	5	ACT2_TOBAC	336	1	P93374	nicotiana t
159	23.8	5	ACT3_LYCES	336	1	Q96484	lycopersico
160	23.8	5	ACT4_LYCES	336	1	Q96481	lycopersico
161	23.8	5	ACT4_TOBAC	336	1	P93372	nicotiana t
162	23.8	5	ACT5_SOLTU	336	1	P81228	solanum tub
163	23.8	5	ACT5_TOBAC	336	1	Q96371	nicotiana t
164	23.8	5	ACT6_TOBAC	336	1	P93376	nicotiana t
165	23.8	5	ACT7_TOBAC	336	1	P93375	nicotiana t
166	23.8	5	ACT9_SOLTU	336	1	P93584	solanum tub
167	23.8	5	ACT8_SOLTU	336	1	P81229	solanum tub
168	23.8	5	AGLF_RHIME	336	1	Q9z3r6	rhizobium m
169	23.8	5	GCB_MOUSE	336	1	P01866	mus musculu
170	23.8	5	A85A_MYCUL	337	1	P58248	mycobacteri
171	23.8	5	ACT4_SOLTU	337	1	P93585	solanum tub
172	23.8	5	YJMC_BACSU	337	1	O34736	bacillus su
173	23.8	5	FLIG_PSEAE	338	1	Q51464	pseudomonas
174	23.8	5	YBD1_YEAST	338	1	P38200	saccharomyc
175	23.8	5	THPS_SULAC	340	1	P17118	sulfolobus
176	23.8	5	VP3_GFLV	341	1	P17768	grapevine f
177	23.8	5	COBT_METSQ	342	1	Q9x7f4	methylobact
178	23.8	5	ILVC_BACSU	342	1	P37253	bacillus su
179	23.8	5	BBUR_BORBR	343	1	O06703	bordetella

Accession	Species	Name	N	Num	Score	Len	Idx
P00795	sus scrofa	CATD_PIG	1	345	23.8	5	180
P55525	rhizobium s	Y4KE_RHISN	1	347	23.8	5	181
Q27556	drosophila	GPDA_DROAE	1	349	23.8	5	182
Q27567	drosophila	GPDA_DROEZ	1	349	23.8	5	183
P58667	clostridium	ISPG_CLOPE	1	349	23.8	5	184
P48439	saccharomyc	OSTG_YEAST	1	350	23.8	5	185
P31561	euglena gra	YCXA_EUGGR	1	350	23.8	5	186
P07735	drosophila	GPDA_DROVI	1	352	23.8	5	187
P17653	marek's dis	KITH_HSVMD	1	352	23.8	5	188
P53491	acetabulari	ACT_ACECL	1	353	23.8	5	189
Q27928	drosophila	GPDA_DROPS	1	353	23.8	5	190
P57203	buchnera ap	HIS7_BUCAI	1	353	23.8	5	191
P46506	caenorhabdi	SB11_CAEEL	1	353	23.8	5	192
P31672	lactobacill	NIFS_LACDE	1	355	23.8	5	193
P53503	oxytricha f	ACT1_OXYFA	1	357	23.8	5	194
P02583	oxytricha f	ACT2_OXYFA	1	357	23.8	5	195
P30172	solanum tub	ACTC_SOLTU	1	357	23.8	5	196
O97463	drosophila	GPDA_DROKA	1	359	23.8	5	197
P13706	drosophila	GPDA_DROME	1	363	23.8	5	198
Q38899	arabidopsis	ORC2_ARATH	1	363	23.8	5	199
P28616	borrelia bu	DNAJ_BORBU	1	364	23.8	5	200
Q9mzs8	ovis aries	CATD_SHEEP	1	365	23.8	5	201
P07012	escherichia	RF2_ECOLI	1	365	23.8	5	202
P43918	haemophilus	RF2_HAEIN	1	365	23.8	5	203
P28353	salmonella	RF2_SALTY	1	365	23.8	5	204
P24263	physarum po	ACTD_PHYPO	1	366	23.8	5	205
P54957	bacillus su	YXER_BACSU	1	370	23.8	5	206
P27132	naegleria f	ACT2_NAEFO	1	371	23.8	5	207
P53458	diphyllobot	ACT5_DIPDE	1	371	23.8	5	208
P73408	synechocyst	YI40_SYNY3	1	372	23.8	5	209
P91754	lumbricus r	ACT_LUMRU	1	372	23.8	5	210
P53499	chondrus cr	ACT_CHOCR	1	373	23.8	5	211
P12431	strongyloce	ACTM_STRPU	1	374	23.8	5	212
P02578	acanthamoeb	ACT1_ACACA	1	375	23.8	5	213
P02577	dictyostell	ACT1_DICDI	1	375	23.8	5	214
P35632	echinococcu	ACT1_ECHGR	1	375	23.8	5	215
P53484	fugu rubrip	ACT1_FUGRU	1	375	23.8	5	216
P02582	zea mays (m	ACT1_MAIZE	1	375	23.8	5	217
P27131	naegleria f	ACT1_NAEFO	1	375	23.8	5	218
P12715	oxytricha n	ACT1_OXYNO	1	375	23.8	5	219
P53468	oxytricha t	ACT1_OXYTR	1	375	23.8	5	220
P43239	pneumocysti	ACT1_PNECA	1	375	23.8	5	221
Q9y702	schizophyll	ACT1_SCHCO	1	375	23.8	5	222
Q9y701	suillus bov	ACT1_SUIBO	1	375	23.8	5	223
P10992	tetrahymena	ACT1_TETTH	1	375	23.8	5	224
P07827	dictyostell	ACT2_DICDI	1	375	23.8	5	225
P53485	fugu rubrip	ACT2_FUGRU	1	375	23.8	5	226
P55805	oxytricha n	ACT2_OXYNO	1	375	23.8	5	227
P53469	oxytricha t	ACT2_OXYTR	1	375	23.8	5	228
Q9y896	schizophyll	ACT2_SCHCO	1	375	23.8	5	229
Q9y707	suillus bov	ACT2_SUIBO	1	375	23.8	5	230
P07829	dictyostell	ACT3_DICDI	1	375	23.8	5	231
P53486	fugu rubrip	ACT3_FUGRU	1	375	23.8	5	232
P07830	dictyostell	ACT8_DICDI	1	375	23.8	5	233
P02576	physarum po	ACTA_PHYPO	1	375	23.8	5	234
P48975	cricetulus	ACTB_CRIGR	1	375	23.8	5	235
P12714	cyprinus ca	ACTB_CYPCA	1	375	23.8	5	236
P02570	homo sapien	ACTB_HUMAN	1	375	23.8	5	237
P29751	oryctolagus	ACTB_RABIT	1	375	23.8	5	238
O42161	salmo salar	ACTB_SALSA	1	375	23.8	5	239
Q93129	branchiosto	ACTC_BRABE	1	375	23.8	5	240
Q93131	branchiosto	ACTC_BRAFL	1	375	23.8	5	241
O17503	branchiosto	ACTC_BRALA	1	375	23.8	5	242
Q00215	styela plic	ACTC_STYPL	1	375	23.8	5	243
Q9uvw9	cephalospor	ACTG_CEPAC	1	375	23.8	5	244
P20359	emericella	ACTG_EMENI	1	375	23.8	5	245
P02571	homo sapien	ACTG_HUMAN	1	375	23.8	5	246
Q9urs0	penicillium	ACTG_PENCH	1	375	23.8	5	247
P53455	ajellomyces	ACT_AJECA	1	375	23.8	5	248
O13419	botrytis ci	ACT_BOTCI	1	375	23.8	5	249
Q9uvx4	coprinus ci	ACT_COPCI	1	375	23.8	5	250
P48465	cryptococcu	ACT_CRYNE	1	375	23.8	5	251
P53502	fucus disti	ACT_FUCDI	1	375	23.8	5	252

#	Length			Name		Organism	Accession
253	375	23.8	5	ACT_NEUCR	1	neurospora	P78711
254	375	23.8	5	ACT_PHARH	1	phaffia rho	P53689
255	375	23.8	5	ACT_PUCGR	1	puccinia gr	P50138
256	375	23.8	5	ACT_SCHPO	1	schizosacch	P10989
257	375	23.8	5	ACT_THELA	1	thermomyces	P10365
258	376	23.8	5	ACT1_AEDAE	1	aedes aegyp	P49128
259	376	23.8	5	ACT1_ARTSX	1	artemia sp.	P18600
260	376	23.8	5	ACT1_BOMMO	1	bombyx mori	P07836
261	376	23.8	5	ACT1_CAEEL	1	caenorhabdi	P10983
262	376	23.8	5	ACT1_DROME	1	drosophila	P10987
263	376	23.8	5	ACT1_HELER	1	heliocidari	P53462
264	376	23.8	5	ACT1_LUMTE	1	lumbricus t	P92182
265	376	23.8	5	ACT1_LYTPI	1	lytechinus	P53465
266	376	23.8	5	ACT1_ONCVO	1	onchocerca	P30162
267	376	23.8	5	ACT1_PEA	1	pisum sativ	P30164
268	376	23.8	5	ACT1_PHYIN	1	phytophthor	P22131
269	376	23.8	5	ACT1_PLAFA	1	plasmodium	P10988
270	376	23.8	5	ACT1_PODCA	1	podocoryne	P41112
271	376	23.8	5	ACT1_SACKO	1	saccoglossu	O18500
272	376	23.8	5	ACT1_SCHMA	1	schistosoma	P53470
273	376	23.8	5	ACT1_STRFN	1	strongyloce	P10990
274	376	23.8	5	ACT2_ARTSX	1	artemia sp.	P18601
275	376	23.8	5	ACT2_BACDO	1	bactrocera	P45885
276	376	23.8	5	ACT2_BOMMO	1	bombyx mori	P07837
277	376	23.8	5	ACT2_CAEEL	1	caenorhabdi	P10984
278	376	23.8	5	ACT2_DIPDE	1	diphyllobot	P53456
279	376	23.8	5	ACT2_DROME	1	drosophila	P02572
280	376	23.8	5	ACT2_ECHGR	1	echinococcu	Q03341
281	376	23.8	5	ACT2_LUMTE	1	lumbricus t	P92176
282	376	23.8	5	ACT2_LYTPI	1	lytechinus	P53466
283	376	23.8	5	ACT2_ONCVO	1	onchocerca	P30163
284	376	23.8	5	ACT2_PEA	1	pisum sativ	P30165
285	376	23.8	5	ACT2_PLAFA	1	plasmodium	P14883
286	376	23.8	5	ACT2_SACKO	1	saccoglossu	O18500
287	376	23.8	5	ACT2_SCHMA	1	schistosoma	P53471
288	376	23.8	5	ACT2_TETPY	1	tetrahymena	P10993
289	376	23.8	5	ACT3_BACDO	1	bactrocera	P45886
290	376	23.8	5	ACT3_BOMMO	1	bombyx mori	P04829
291	376	23.8	5	ACT3_DROME	1	drosophila	P53501
292	376	23.8	5	ACT3_HELAM	1	helicoverpa	Q25010
293	376	23.8	5	ACT3_LIMPO	1	limulus pol	P41340
294	376	23.8	5	ACT3_ORYSA	1	oryza sativ	P17299
295	376	23.8	5	ACT3_PODCA	1	podocoryne	P41113
296	376	23.8	5	ACT3_SOYBN	1	glycine max	P02580
297	376	23.8	5	ACT4_ARTSX	1	artemia sp.	P18603
298	376	23.8	5	ACT4_BOMMO	1	bombyx mori	Q27250
299	376	23.8	5	ACT4_CAEEL	1	caenorhabdi	P10986
300	376	23.8	5	ACT4_DROME	1	drosophila	P02574
301	376	23.8	5	ACT5_BACDO	1	bactrocera	P45887
302	376	23.8	5	ACT5_CHICK	1	gallus gall	P53478
303	376	23.8	5	ACT5_DROME	1	drosophila	P10981
304	376	23.8	5	ACT5_XENLA	1	xenopus lae	P53505
305	376	23.8	5	ACT6_DROME	1	drosophila	P02575
306	376	23.8	5	ACT7_ORYSA	1	oryza sativ	P17300
307	376	23.8	5	ACT8_XENLA	1	xenopus lae	P53506
308	376	23.8	5	ACTA_LIMPO	1	limulus pol	P41339
309	376	23.8	5	ACTA_STRPU	1	strongyloce	P53472
310	376	23.8	5	ACTB_STRPU	1	strongyloce	P53473
311	376	23.8	5	ACTC_HALRO	1	halocynthia	P53461
312	376	23.8	5	ACTC_PISOC	1	pisaster oc	P12716
313	376	23.8	5	ACTD_STRPU	1	strongyloce	Q07903
314	376	23.8	5	ACTE_STRPU	1	strongyloce	P10991
315	376	23.8	5	ACTF_STRPU	1	strongyloce	P53474
316	376	23.8	5	ACTH_HUMAN	1	homo sapien	P18499
317	376	23.8	5	ACTM_APLCA	1	aplysia cal	P12718
318	376	23.8	5	ACTM_HELER	1	heliocidari	P17304
319	376	23.8	5	ACTM_HELTB	1	heliocidari	P53463
320	376	23.8	5	ACTM_PISOC	1	pisaster oc	P53464
321	376	23.8	5	ACTM_PISOC	1	pisaster oc	P12717
322	376	23.8	5	ACTX_FUGRU	1	fugu rubrip	P53483
323	376	23.8	5	ACTY_LIMPO	1	limulus pol	P41341
324	376	23.8	5	ACT_ACHBI	1	achlya bise	P26182
325	376	23.8	5	ACT_BIOGL	1	biomphalari	P92179

#	Accession	Organism	Name		Length		
326	P90689	brugia mala	ACT_BRUMA	1	376	23.8	5
327	P14235	candida alb	ACT_CANAL	1	376	23.8	5
328	Q9uvz8	candida dub	ACT_CANDU	1	376	23.8	5
329	O17320	crassostrea	ACT_CRAGI	1	376	23.8	5
330	P26183	cryptospori	ACT_CRYPV	1	376	23.8	5
331	P11426	entamoeba h	ACT_ENTHI	1	376	23.8	5
332	Q39758	fucus vesic	ACT_FUCVE	1	376	23.8	5
333	P17126	hydra atten	ACT_HYDAT	1	376	23.8	5
334	P49871	manduca sex	ACT_MANSE	1	376	23.8	5
335	O16808	mayetiola d	ACT_MAYDE	1	376	23.8	5
336	O74258	pichia angu	ACT_PICAN	1	376	23.8	5
337	Q26065	placopecten	ACT_PLAMG	1	376	23.8	5
338	P14227	taenia soli	ACT_TAESO	1	376	23.8	5
339	P53476	toxoplasma	ACT_TOXGO	1	376	23.8	5
340	O43502	homo sapien	R51C_HUMAN	1	376	23.8	5
341	P49942	bacteroides	XYNA_BACOV	1	376	23.8	5
342	Q98972	oryzias lat	ACT1_ORYLA	1	377	23.8	5
343	P13362	oryza sativ	ACT1_ORYSA	1	377	23.8	5
344	P53504	sorghum bic	ACT1_SORBI	1	377	23.8	5
345	P02581	glycine max	ACT1_SOYBN	1	377	23.8	5
346	Q05214	nicotiana t	ACT1_TOBAC	1	377	23.8	5
347	P04751	xenopus lae	ACT1_XENLA	1	377	23.8	5
348	P26197	absidia gla	ACT2_ABSGL	1	377	23.8	5
349	P53492	arabidopsis	ACT2_ARATH	1	377	23.8	5
350	P10995	xenopus lae	ACT2_XENLA	1	377	23.8	5
351	P20399	xenopus tro	ACT2_XENTR	1	377	23.8	5
352	P53493	arabidopsis	ACT3_ARATH	1	377	23.8	5
353	P53457	diphyllobot	ACT3_DIPDE	1	377	23.8	5
354	P46258	pisum sativ	ACT3_PEA	1	377	23.8	5
355	P30167	solanum tub	ACT3_SOLTU	1	377	23.8	5
356	P04752	xenopus lae	ACT3_XENLA	1	377	23.8	5
357	P53494	arabidopsis	ACT4_ARATH	1	377	23.8	5
358	P30168	solanum tub	ACT6_SOLTU	1	377	23.8	5
359	P30169	solanum tub	ACT7_SOLTU	1	377	23.8	5
360	P08023	gallus gall	ACTA_CHICK	1	377	23.8	5
361	P03996	homo sapien	ACTA_HUMAN	1	377	23.8	5
362	P53496	arabidopsis	ACTB_ARATH	1	377	23.8	5
363	P30171	solanum tub	ACTB_SOLTU	1	377	23.8	5
364	P53497	arabidopsis	ACTC_ARATH	1	377	23.8	5
365	P53480	fugu rubrip	ACTC_FUGRU	1	377	23.8	5
366	P04270	homo sapien	ACTC_HUMAN	1	377	23.8	5
367	P30173	solanum tub	ACTD_SOLTU	1	377	23.8	5
368	P49055	carassius a	ACTS_CARAU	1	377	23.8	5
369	P53479	cyprinus ca	ACTS_CYPCA	1	377	23.8	5
370	P53481	fugu rubrip	ACTS_FUGRU	1	377	23.8	5
371	P02568	homo sapien	ACTS_HUMAN	1	377	23.8	5
372	P53482	fugu rubrip	ACTT_FUGRU	1	377	23.8	5
373	P53498	chlamydomon	ACT_CHLRE	1	377	23.8	5
374	O65315	coleochaete	ACT_COLSC	1	377	23.8	5
375	P53500	cyanidiosch	ACT_CYAME	1	377	23.8	5
376	O65316	mesostigma	ACT_MESVI	1	377	23.8	5
377	P20904	volvox cart	ACT_VOLCA	1	378	23.8	5
378	P53460	halocynthia	ACT1_HALRO	1	378	23.8	5
379	P27130	halocynthia	ACT2_HALRO	1	378	23.8	5
380	Q25472	molgula ocu	ACT2_MOLOC	1	378	23.8	5
381	Q93130	branchiosto	ACTM_BRAFL	1	378	23.8	5
382	O15998	ciona savig	ACTM_CIOSA	1	378	23.8	5
383	P53467	molgula ocu	ACTM_MOLOC	1	378	23.8	5
384	P26198	styela clav	ACTM_STYCL	1	378	23.8	5
385	P53475	styela clav	ACTN_STYCL	1	378	23.8	5
386	O65314	scheffelia	ACT1_SCHDU	1	378	23.8	5
387	P17298	oryza sativ	ACT2_ORYSA	1	379	23.8	5
388	Q93130	branchiosto	ACTM_BRABE	1	379	23.8	5
389	O00214	styela plic	ACTM_STYPL	1	379	23.8	5
390	P22343	daucus caro	ACT1_DAUCA	1	380	23.8	5
391	O17502	branchiosto	ACTM_BRALA	1	380	23.8	5
392	O92mw7	helicobacte	METB_HELPJ	1	380	23.8	5
393	P55069	helicobacte	METB_HELPY	1	380	23.8	5
394	Q00374	coletotric	PLYA_COLGL	1	380	23.8	5
395	P23344	daucus caro	ACT2_DAUCA	1	381	23.8	5
396	P42023	pneumocysti	ACTZ_PNECA	1	385	23.8	5
397	Q03168	aedes aegyp	ASPP_AEDAE	1	387	23.8	5
398	P35515	methanosarc	DNAJ_METMA	1	389	23.8	5

Index	Score	Len	ID	Flag	Name	Accession	Organism
399	23.8	5	390	1	CATD_BOVIN	P80209	bos taurus
400	23.8	5	390	1	RRPP_MUMP1	P19717	mumps virus
401	23.8	5	390	1	Y4R1_RHISN	P55642	rhizobium s
402	23.8	5	391	1	GAL1_MOUSE	Q9r0n0	mus musculu
403	23.8	5	391	1	RRPP_MUMPE	P16072	mumps virus
404	23.8	5	391	1	RRPP_MUMPM	P16595	mumps virus
405	23.8	5	392	1	PAF2_BOVIN	P79106	bos taurus
406	23.8	5	392	1	PAF2_HUMAN	Q99487	homo sapien
407	23.8	5	392	1	SB11_HUMAN	Q96pl5	homo sapien
408	23.8	5	393	1	GC1M_MOUSE	P01869	mus musculu
409	23.8	5	394	1	CIW3_HUMAN	O14649	homo sapien
410	23.8	5	395	1	NEUA_STRAG	Q53598	streptococc
411	23.8	5	396	1	BRB2_RAT	P25003	rattus norv
412	23.8	5	396	1	CARP_NEUCR	Q01394	neurospora
413	23.8	5	396	1	SOTB_HAEIN	P44535	haemophilus
414	23.8	5	396	1	YD18_YEAST	Q12185	saccharomyc
415	23.8	5	396	1	YWBD_BACSU	P39587	bacillus su
416	23.8	5	397	1	6P22_YEAST	Q12471	saccharomyc
417	23.8	5	397	1	LHX3_HUMAN	Q9ubr4	homo sapien
418	23.8	5	398	1	APL1_HUMAN	O14791	homo sapien
419	23.8	5	398	1	TRA5_RHIME	Q52873	rhizobium m
420	23.8	5	399	1	GCAM_MOUSE	P01865	mus musculu
421	23.8	5	400	1	PAI1_MUSVI	P50449	mustela vis
422	23.8	5	402	1	V063_FOWPV	P21971	fowlpox vir
423	23.8	5	402	1	PAI1_BOVIN	P13909	bos taurus
424	23.8	5	402	1	PAI1_HUMAN	P05121	homo sapien
425	23.8	5	402	1	PAI1_PIG	P79335	sus scrofa
426	23.8	5	402	1	PGK_CHLPN	Q9z7m5	chlamydia p
427	23.8	5	405	1	GCBM_MOUSE	P01867	mus musculu
428	23.8	5	407	1	CATD_RAT	P24268	rattus norv
429	23.8	5	408	1	ENV_RSFFV	P03389	rauscher sp
430	23.8	5	409	1	RUN3_MOUSE	Q64131	m runt-rela
431	23.8	5	410	1	CATD_MOUSE	P18242	mus musculu
432	23.8	5	411	1	Y360_MYCGE	Q49426	mycoplasma
433	23.8	5	412	1	CATD_HUMAN	P07339	homo sapien
434	23.8	5	414	1	PYRC_THEAC	Q9himO	thermoplasm
435	23.8	5	415	1	RUN3_HUMAN	Q13761	h runt-rela
436	23.8	5	416	1	NH59_CAEEL	Q9txj1	caenorhabdi
437	23.8	5	416	1	TRA1_HUMAN	Q13077	homo sapien
438	23.8	5	417	1	EX71_LACLA	Q9ch84	lactococcus
439	23.8	5	418	1	RFBE_SHIFL	P37781	shigella fl
440	23.8	5	419	1	HAT1_HUMAN	O14929	homo sapien
441	23.8	5	419	1	NIFS_FRASE	Q9z5x5	frankia sp.
442	23.8	5	421	1	EXOF_RHIME	Q02728	rhizobium m
443	23.8	5	421	1	HMDH_AERPE	Q9yas4	aeropyrum p
444	23.8	5	423	1	SYY_ECOLI	P00951	escherichia
445	23.8	5	427	1	IM44_SCHPO	O60084	schizosacch
446	23.8	5	428	1	YFJD_ECOLI	P37908	escherichia
447	23.8	5	432	1	ENV2_MOUSE	P11370	mus musculu
448	23.8	5	432	1	PTN1_MOUSE	P35821	mus musculu
449	23.8	5	432	1	PTN1_RAT	P20417	rattus norv
450	23.8	5	434	1	YTFL_HAEIN	P44717	haemophilus
451	23.8	5	434	1	CBPS_STRCP	P39041	streptomyce
452	23.8	5	435	1	MALC_STRPN	Q04698	streptococc
453	23.8	5	435	1	PTN1_HUMAN	P18031	homo sapien
454	23.8	5	436	1	EPE1_RALSO	P58594	ralstonia s
455	23.8	5	436	1	EPE2_RALSO	Q45411	ralstonia s
456	23.8	5	437	1	ADFP_HUMAN	Q99541	homo sapien
457	23.8	5	440	1	BCHN_HELMO	Q9zge9	heliobacill
458	23.8	5	442	1	SSNA_ECOLI	Q46812	escherichia
459	23.8	5	443	1	FLVS_FLAME	Q47899	flavobacter
460	23.8	5	443	1	Y082_SYNY3	Q55803	synechocyst
461	23.8	5	444	1	CADB_ECOLI	P23891	escherichia
462	23.8	5	449	1	PUR1_PYRHO	O57979	pyrococcus
463	23.8	5	450	1	ADFP_BOVIN	Q9tun6	bos taurus
464	23.8	5	451	1	TBG_ENTHI	P54401	entamoeba h
465	23.8	5	451	1	VU10_HSV7J	P52524	human herpe
466	23.8	5	454	1	VNUC_THOGV	P89216	thogoto vir
467	23.8	5	457	1	IRF7_MOUSE	P70434	mus musculu
468	23.8	5	461	1	SSL1_YEAST	Q04673	saccharomyc
469	23.8	5	463	1	VL2_HPV13	Q02275	human papil
470	23.8	5	464	1	VNSS_TSWV1	P26002	tomato spot
471	23.8	5	467	1	ATPX_BACFI	P25075	bacillus fi
472	23.8	5	467	1	CBPA_DICDI	P35085	dictyosteli
473	23.8	5	467	1	VNSS_TSWVL	P26003	tomato spot
474	23.8	5	468	1	T10A_HUMAN	O00220	homo sapien
475	23.8	5	471	1	IF2G_HUMAN	P41091	homo sapien
476	23.8	5	471	1	IF2G_MOUSE	Q9z0n1	mus musculu
477	23.8	5	473	1	IF2H_MOUSE	Q9z0n2	mus musculu
478	23.8	5	473	1	CB1R_POEGU	P56971	poephila gu
479	23.8	5	474	1	GATA_SCHPO	O13837	schizosacch
480	23.8	5	474	1	PRTP_MOUSE	P16675	mus musculu
481	23.8	5	475	1	IF26_DROME	Q24208	drosophila
482	23.8	5	478	1	OPN4_HUMAN	Q9uhm6	homo sapien
483	23.8	5	479	1	ORC5_YEAST	P50874	saccharomyc
484	23.8	5	480	1	PRTP_HUMAN	P10619	homo sapien
485	23.8	5	480	1	Y523_XYLFA	O68639	xylella fas
486	23.8	5	482	1	TISD_HUMAN	P47974	homo sapien
487	23.8	5	484	1	NTRC_RHIME	P10577	rhizobium m
488	23.8	5	485	1	N070_SOYBN	Q02920	glycine max
489	23.8	5	491	1	PBP_BACSU	P39844	bacillus su
490	23.8	5	492	1	CAT2_RICCO	P49318	ricinus com
491	23.8	5	498	1	GATA_EMENI	P14010	emericella
492	23.8	5	498	1	MEFD_XENLA	O03413	xenopus lae
493	23.8	5	498	1	YDHP_SCHPO	Q92360	schizosacch
494	23.8	5	501	1	LON2_METTH	O26978	methanobact
495	23.8	5	505	1	MAL0_SYNY3	P72785	synechocyst
496	23.8	5	506	1	2157_HUMAN	P51786	homo sapien
497	23.8	5	507	1	IRX3_MOUSE	P81067	mus musculu
498	23.8	5	508	1	MKK1_YEAST	P32490	saccharomyc
499	23.8	5	508	1	VL1_HPV1A	P03099	human papil
500	23.8	5	509	1	FLAA_HELPY	Q03843	helicobacte
501	23.8	5	510	1	PRO1_LISMO	P23224	listeria mo
502	23.8	5	510	1	PRO2_LISMO	P34025	listeria mo
503	23.8	5	512	1	AMY1_DEBOC	P19269	debaryomyce
504	23.8	5	512	1	FLAA_HELFE	Q9xb38	helicobacte
505	23.8	5	514	1	TSSP_HUMAN	Q9nqe7	homo sapien
506	23.8	5	517	1	DHAS_HUMAN	P30837	homo sapien
507	23.8	5	518	1	RN23_HUMAN	Q9hcm9	homo sapien
508	23.8	5	521	1	VENV_DHVI1	P27427	dhori virus
509	23.8	5	523	1	GDS1_YEAST	P41913	saccharomyc
510	23.8	5	523	1	RF3_LACLA	Q9c1k7	lactococcus
511	23.8	5	524	1	LCYE_ARATH	Q38932	arabidopsis
512	23.8	5	526	1	C4E3_DROME	Q9v192	drosophila
513	23.8	5	536	1	YRN3_CAEEL	Q09606	caenorhabdi
514	23.8	5	541	1	PELP_YERPS	P11278	yersinia ps
515	23.8	5	543	1	APPA_BACSU	P42061	bacillus su
516	23.8	5	545	1	YB90_YEAST	P38346	saccharomyc
517	23.8	5	553	1	SECY_MAIZE	O63066	zea mays (m
518	23.8	5	556	1	INX7_CAEEL	Q21123	caenorhabdi
519	23.8	5	564	1	HEMA_IADCZ	P19696	influenza a
520	23.8	5	564	1	HEMA_IAGRE	P19698	influenza a
521	23.8	5	568	1	PELP_ERWCA	P14005	erwinia car
522	23.8	5	571	1	PRIM_DEIRA	Q9rwr5	deinococcus
523	23.8	5	578	1	YTFM_HAEIN	P44038	haemophilus
524	23.8	5	581	1	YG25_ARCFU	O28648	archaeoglob
525	23.8	5	585	1	YW04_SCHPO	Q9hgm7	schizosacch
526	23.8	5	586	1	CYDD_HAEIN	P45082	haemophilus
527	23.8	5	586	1	KPYK_BACPY	P51182	bacillus ps
528	23.8	5	586	1	RRPO_BWYVF	P09507	beet wester
529	23.8	5	587	1	CYA3_RHIME	Q9zgq0	rhizobium m
530	23.8	5	589	1	YE67_HAEIN	P45221	haemophilus
531	23.8	5	590	1	PEPF_BORBU	O51264	borrelia bu
532	23.8	5	593	1	SLT_HAEIN	P44888	haemophilus
533	23.8	5	593	1	HM21_HUMAN	Q9ugu5	homo sapien
534	23.8	5	599	1	LAM2_HUMAN	Q03252	homo sapien
535	23.8	5	599	1	YOU3_CAEEL	P30639	caenorhabdi
536	23.8	5	608	1	GHR_CHICK	Q02092	gallus gall
537	23.8	5	610	1	RHO_MYCLE	P45835	mycobacteri
538	23.8	5	614	1	RPB1_METVA	P41557	methanococc
539	23.8	5	614	1	YQZ2_CAEEL	Q09320	caenorhabdi
540	23.8	5	617	1	41_BOVIN	Q9n179	bos taurus
541	23.8	5	618	1	VE1_HPV63	Q07847	human papil
542	23.8	5	621	1	PLAK_MOUSE	Q02257	mus musculu
543	23.8	5	625	1	AMYG_NEUCR	P14804	neurospora
544	23.8	5	636	1	ENV_MCFF	P15073	mink cell f

No.			Name		Len	Accession	Organism
545	23.8	5	RPB1_METJA	1	636	Q60181	methanococc
546	23.8	5	YD14_SCHPO	1	638	Q10237	schizosacch
547	23.8	5	ENV_RMCFV	1	640	P06445	rauscher ml
548	23.8	5	SYT_THEMA	1	640	Q9wzj9	thermotoga
549	23.8	5	FIB2_PETMA	1	641	P33573	petromyzon
550	23.8	5	IMD_ARTGO	1	641	Q44052	arthrobacte
551	23.8	5	ARS_CHLRE	1	647	P14217	chlamydomon
552	23.8	5	PPE1_HUMAN	1	653	O14829	homo sapien
553	23.8	5	YC26_PORPU	1	656	P51392	porphyra pu
554	23.8	5	CPT2_HUMAN	1	658	P23786	homo sapien
555	23.8	5	CPT2_MOUSE	1	658	P52825	mus musculu
556	23.8	5	CPT2_RAT	1	658	P18886	rattus norv
557	23.8	5	ENV_MLVCB	1	661	P08360	cas-br-e mu
558	23.8	5	Y4ID_RHISN	1	662	P55487	rhizobium s
559	23.8	5	ENV_MLVMO	1	665	P03385	moloney mur
560	23.8	5	ENV_MLVHO	1	666	P21436	homulv muri
561	23.8	5	YCDR_ECOLI	1	672	P75906	escherichia
562	23.8	5	ENV_MLVF5	1	675	P03390	friend muri
563	23.8	5	NED1_MOUSE	1	675	P33215	mus musculu
564	23.8	5	ENV_MLVFP	1	676	P26803	friend muri
565	23.8	5	TRF_MANSE	1	681	P22297	manduca sex
566	23.8	5	PRC_SALTY	1	682	P43669	salmonella
567	23.8	5	SNK_MOUSE	1	682	P53351	mus musculu
568	23.8	5	SNK_RAT	1	682	Q9r012	rattus norv
569	23.8	5	SNK_HUMAN	1	685	Q9nyy3	homo sapien
570	23.8	5	SNWA_DICDI	1	685	P54705	dictyosteli
571	23.8	5	EOMD_HUMAN	1	686	O95936	homo sapien
572	23.8	5	KLC_STRPU	1	686	Q05090	strongyloce
573	23.8	5	DNLJ_RICPR	1	689	Q9zck9	rickettsia
574	23.8	5	VG42_BPMU	1	690	Q9tlv6	bacteriopha
575	23.8	5	YDHV_ECOLI	1	700	P76192	escherichia
576	23.8	5	YBY1_YEAST	1	704	P38273	saccharomyc
577	23.8	5	FRE2_YEAST	1	711	P36033	saccharomyc
578	23.8	5	CYAB_BORPE	1	712	P18770	bordetella
579	23.8	5	ATI2_HSV1F	1	715	P08314	herpes simp
580	23.8	5	ATI2_HSV11	1	718	P10230	herpes simp
581	23.8	5	MUS2_STAAU	1	719	Q9zeh5	staphylococ
582	23.8	5	EF2_METBU	1	730	O93632	methanococc
583	23.8	5	EF2_METMT	1	730	O93637	methanococc
584	23.8	5	CNG1_CHICK	1	735	Q00805	gallus gall
585	23.8	5	MCM5_XENLA	1	735	P55862	xenopus lae
586	23.8	5	PLAK_XENLA	1	738	P30998	xenopus lae
587	23.8	5	KM65_YEAST	1	742	Q03656	saccharomyc
588	23.8	5	PLAK_HUMAN	1	743	P14923	homo sapien
589	23.8	5	MUS2_THEMA	1	757	Q9x105	thermotoga
590	23.8	5	PHSA_SALTY	1	758	P37600	salmonella
591	23.8	5	DAB2_RAT	1	768	O88797	rattus norv
592	23.8	5	PAC2_PSES3	1	773	P15558	pseudomonas
593	23.8	5	STF_LAMBD	1	774	P03764	bacteriopha
594	23.8	5	MUS2_BORBU	1	780	O51125	borrelia bu
595	23.8	5	MUS2_BACHD	1	785	Q9k8a0	bacillus ha
596	23.8	5	MUS2_BACSU	1	785	P94545	bacillus su
597	23.8	5	YHY2_YEAST	1	785	P38870	saccharomyc
598	23.8	5	UGS3_SOLTU	1	788	Q43847	solanum tub
599	23.8	5	SYFB_CHLMU	1	790	Q9pjr8	chlamydia m
600	23.8	5	RSG4_MOUSE	1	799	Q92268	mus musculu
601	23.8	5	41_XENLA	1	801	P11434	xenopus lae
602	23.8	5	PEFC_SALTY	1	802	P37868	salmonella
603	23.8	5	FAED_ECOLI	1	812	P06970	escherichia
604	23.8	5	AKH_HAEIN	1	815	P44505	haemophilus
605	23.8	5	AKH_BUCAI	1	816	P57290	buchnera ap
606	23.8	5	AK1H_SERMA	1	819	P27725	serratia ma
607	23.8	5	AK1H_ECOLI	1	820	P00561	escherichia
608	23.8	5	JIP2_HUMAN	1	824	Q13387	homo sapien
609	23.8	5	JIP2_MOUSE	1	829	P22223	homo sapien
610	23.8	5	CAD3_HUMAN	1	830	Q9ere9	mus musculu
611	23.8	5	OBP_VZVD	1	835	P09299	varicella-z
612	23.8	5	DEGX_CAEEL	1	849	Q09274	caenorhabdi
613	23.8	5	OBP_HSV11	1	851	P10193	herpes simp
614	23.8	5	41_MOUSE	1	858	P48193	mus musculu
615	23.8	5	M3KC_HUMAN	1	859	Q12852	homo sapien
616	23.8	5	4.1_HUMAN	1	864	P11171	homo sapien
617	23.8	5	PQP1_YEAST	1	875	P41812	saccharomyc

No.			Name		Len	Accession	Organism
618	23.8	5	E4L1_MOUSE	1	879	Q92h5	mus musculu
619	23.8	5	YDBH_ECOLI	1	879	P52645	escherichia
620	23.8	5	E4L1_HUMAN	1	881	Q9h4g0	homo sapien
621	23.8	5	M3KC_MOUSE	1	888	Q60790	mus musculu
622	23.8	5	M3KC_RAT	1	888	O63796	rattus norv
623	23.8	5	RRN6_YEAST	1	894	P32786	saccharomyc
624	23.8	5	VP3_EHDV1	1	899	P27281	epizootic h
625	23.8	5	VP3_EHDVA	1	899	P33474	epizootic h
626	23.8	5	VP3_BTV10	1	901	P12435	bluetongue
627	23.8	5	VP3_BTV11	1	901	Q65749	bluetongue
628	23.8	5	VP3_BTV13	1	901	Q65750	bluetongue
629	23.8	5	VP3_BTV17	1	901	P03539	bluetongue
630	23.8	5	VP3_BTV1A	1	901	P20608	bluetongue
631	23.8	5	VP3_BTV1S	1	901	P56582	bluetongue
632	23.8	5	VP3_BTV2A	1	901	Q65748	bluetongue
633	23.8	5	RIR1_RCMVA	1	906	P16782	human cytom
634	23.8	5	DPO1_BORBU	1	908	Q51498	borrelia bu
635	23.8	5	DLG1_RAT	1	911	Q62696	rattus norv
636	23.8	5	SECA_TREPA	1	916	O83394	treponema p
637	23.8	5	SLAP_THETH	1	917	Q51830	thermus aqu
638	23.8	5	STE5_YEAST	1	917	P32917	saccharomyc
639	23.8	5	PTN4_HUMAN	1	926	P29074	homo sapien
640	23.8	5	PMP9_CHLPN	1	928	Q9z398	chlamydia p
641	23.8	5	E4L3_MOUSE	1	929	Q9wv92	mus musculu
642	23.8	5	PMP8_CHLPN	1	930	Q9z393	chlamydia p
643	23.8	5	PKL1_HUMAN	1	942	Q16512	homo sapien
644	23.8	5	YM11_YEAST	1	943	P39523	saccharomyc
645	23.8	5	PKL1_RAT	1	946	Q63433	rattus norv
646	23.8	5	UVRA_BORBU	1	950	O51777	borrelia bu
647	23.8	5	MTN2_HUMAN	1	956	O00339	homo sapien
648	23.8	5	MTN2_MOUSE	1	956	O08746	mus musculu
649	23.8	5	YEF3_YEAST	1	957	P32618	saccharomyc
650	23.8	5	AMPE_HUMAN	1	957	Q07075	homo sapien
651	23.8	5	MML2_MYCTU	1	968	Q11171	mycobacteri
652	23.8	5	CSE1_DROME	1	975	Q9xzul	drosophila
653	23.8	5	YAA2_SCHPO	1	990	Q09796	schizosacch
654	23.8	5	SCA4_RICSI	1	991	Q9aj77	rickettsia
655	23.8	5	EVC_HUMAN	1	992	P57679	homo sapien
656	23.8	5	Y456_CHLTR	1	1005	O84462	chlamydia t
657	23.8	5	SCA4_RICPA	1	1010	Q9aj75	rickettsia
658	23.8	5	SCA4_RICAF	1	1011	Q9aj83	rickettsia
659	23.8	5	SCA4_RICMN	1	1011	Q9aj82	rickettsia
660	23.8	5	SCA4_RICSL	1	1012	Q9aj80	rickettsia
661	23.8	5	UBA1_SCHPO	1	1012	O94609	schizosacch
662	23.8	5	SCA4_RICRH	1	1013	Q9aj81	rickettsia
663	23.8	5	SCA4_RICJA	1	1018	Q9aj79	rickettsia
664	23.8	5	YKC2_CAEEL	1	1018	P41993	caenorhabdi
665	23.8	5	SCA4_RICCN	1	1022	Q52658	rickettsia
666	23.8	5	PTP1_CAEEL	1	1026	P28191	caenorhabdi
667	23.8	5	SGG_DROME	1	1067	P18431	drosophila
668	23.8	5	E4L1_HUMAN	1	1087	Q9y2j2	homo sapien
669	23.8	5	P14K_DICDI	1	1093	P54677	dictyosteli
670	23.8	5	EMBA_MYCTU	1	1094	Q72060	mycobacteri
671	23.8	5	EX5C_ECOLI	1	1122	P07648	escherichia
672	23.8	5	RPOB_THECE	1	1122	P31814	thermococcu
673	23.8	5	NGAP_HUMAN	1	1139	Q9uf2	homo sapien
674	23.8	5	SRE2_HUMAN	1	1141	Q12772	homo sapien
675	23.8	5	CCAS_RAT	1	1146	Q02485	rattus norv
676	23.8	5	TFC3_YEAST	1	1160	P34111	saccharomyc
677	23.8	5	RPOB_THEAC	1	1195	Q03587	thermoplasm
678	23.8	5	NIFJ_ANASP	1	1199	Q06879	anabaena sp
679	23.8	5	C1BE_BACTU	1	1227	O85805	bacillus th
680	23.8	5	MYS9_DROME	1	1253	Q01989	drosophila
681	23.8	5	GNRP_MOUSE	1	1262	P27671	mus musculu
682	23.8	5	VL3_REOVD	1	1267	P17378	reovirus (t
683	23.8	5	VL3_REOVJ	1	1267	P17377	reovirus (t
684	23.8	5	VL3_REOVL	1	1267	P17376	reovirus (t
685	23.8	5	MDR1_CRIGR	1	1276	P21448	cricetulus
686	23.8	5	MDR1_MOUSE	1	1276	P06795	mus musculu
687	23.8	5	MDR2_CRIGR	1	1276	P21449	cricetulus
688	23.8	5	MDR1_HUMAN	1	1280	P08183	homo sapien
689	23.8	5	RPOC_MYCGE	1	1292	P47582	mycoplasma
690	23.8	5	CA1E_HUMAN	1	1388	P39059	homo sapien

#	n	Tm	Len	col	Name	Accession	Organism
691	5	23.8	1395	1	SP41_YEAST	P38904	saccharomyc
692	5	23.8	1398	1	DNA2_SCHPO	Q9uru2	schizosacch
693	5	23.8	1403	1	BIRE_MOUSE	Q9r016	mus musculu
694	5	23.8	1407	1	RPOC_ECOLI	P00577	escherichia
695	5	23.8	1407	1	RPOC_SALTY	Q91937	salmonella
696	5	23.8	1434	1	PTC1_MOUSE	Q61115	mus musculu
697	5	23.8	1490	1	CRK7_HUMAN	Q9ny4	homo sapien
698	5	23.8	1522	1	PST1_SCHPO	Q09750	schizosacch
699	5	23.8	1557	1	LMG1_CAEEL	Q18823	caenorhabdi
700	5	23.8	1581	1	PPRB_HUMAN	Q15648	h peroxisom
701	5	23.8	1607	1	LMG1_MOUSE	P02468	mus musculu
702	5	23.8	1609	1	LMG1_HUMAN	P11047	homo sapien
703	5	23.8	1687	1	CCAM_MUSDO	Q25452	musca domes
704	5	23.8	1698	1	Y076_HUMAN	Q14999	homo sapien
705	5	23.8	1699	1	POLN_LORDV	P54634	lordsdale v
706	5	23.8	1770	1	RI15_YEAST	P43565	saccharomyc
707	5	23.8	1786	1	YCF1_ARATH	P56785	arabidopsis
708	5	23.8	1818	1	2294_HUMAN	O94822	homo sapien
709	5	23.8	1828	1	CUT1_SCHPO	P18296	schizosacch
710	5	23.8	1856	1	GBF1_CRIGR	Q9r1d7	cricetulus
711	5	23.8	1859	1	GBF1_HUMAN	Q92538	homo sapien
712	5	23.8	1873	1	CCAS_HUMAN	Q13698	homo sapien
713	5	23.8	1873	1	CCAS_RABIT	P07293	oryctolagus
714	5	23.8	1901	1	YCF1_TOBAC	P12222	nicotiana t
715	5	23.8	1951	1	CIN3_HUMAN	Q9ny46	homo sapien
716	5	23.8	1951	1	CIN3_RAT	P08104	rattus norv
717	5	23.8	1959	1	AGR1_RAT	P25304	rattus norv
718	5	23.8	1971	1	MC3A_MOUSE	Q9wuu9	mus musculu
719	5	23.8	2124	1	Y192_HUMAN	Q93074	homo sapien
720	5	23.8	2145	1	CYAA_PODAN	Q01513	podospora a
721	5	23.8	2151	1	RRPL_HANTV	Q23456	hantaan vir
722	5	23.8	2151	1	RRPL_SEOU8	P27314	seoul virus
723	5	23.8	2156	1	RRPL_PUUMH	P27176	puumala vir
724	5	23.8	2223	1	CCAE_DISOM	P56699	discopyge o
725	5	23.8	2248	1	Y539_HUMAN	O60287	homo sapien
726	5	23.8	2254	1	CCAG_RAT	O54898	rattus norv
727	5	23.8	2255	1	RRPL_SV5	Q88434	simian viru
728	5	23.8	2255	1	RRPL_SV5WR	Q03336	simian viru
729	5	23.8	2261	1	RRPL_MUMPM	P30929	mumps virus
730	5	23.8	2262	1	RRPL_PI2HT	P26676	human paral
731	5	23.8	2269	1	RRPL_SV41	P35341	simian viru
732	5	23.8	2273	1	ABCR_HUMAN	P78363	homo sapien
733	5	23.8	2441	1	CBP_MOUSE	Q24270	drosophila
734	5	23.8	2516	1	CCAD_DROME	Q9y618	h nuclear r
735	5	23.8	2517	1	NCR2_HUMAN	P39847	bacillus su
736	5	23.8	2555	1	PPS3_BACSU	P15924	homo sapien
737	5	23.8	2871	1	DESP_HUMAN	P51111	rattus norv
738	5	23.8	3110	1	HD_RAT	P42859	mus musculu
739	5	23.8	3119	1	HD_MOUSE	Q61493	mus musculu
740	5	23.8	3122	1	DPOZ_MOUSE	P17766	p genome po
741	5	23.8	3125	1	POLG_PPVNA	P17767	p genome po
742	5	23.8	3140	1	POLG_PPVRA	Q84934	p genome po
743	5	23.8	3140	1	POLG_PPVSK	P13529	p genome po
744	5	23.8	3141	1	POLG_PPVD	P42858	homo sapien
745	5	23.8	3144	1	HD_HUMAN	P51112	fugu rubrip
746	5	23.8	3148	1	HD_FUGRU	Q02597	t genome po
747	5	23.8	3163	1	POLG_TUMVQ	P89509	t genome po
748	5	23.8	3164	1	POLG_TUMVJ	Q9nrc6	homo sapien
749	5	23.8	3674	1	SPCR_HUMAN	P36022	saccharomyc
750	5	23.8	4092	1	DYHC_YEAST	O13290	schizosacch
751	5	23.8	4196	1	DYHC_SCHPO	P33450	drosophila
752	4	23.8	5147	1	FAT_DROME	Q01886	cochliobolu
753	4	23.8	5217	1	HTS1_COCCA	Q9qxz0	mus musculu
754	4	23.8	5327	1	ACF1_MOUSE	Q9upn3	homo sapien
755	4	23.8	5430	1	ACF7_HUMAN	P38495	callinectes
756	4	19.0	9	1	FARP_CALSI	P80863	bacillus su
757	4	19.0	12	1	SO15_BACSU	P82177	shewanella
758	4	19.0	19	1	MDH_SHEPU	P81682	staphylococ
759	4	19.0	30	1	OTCX_STAEP	P24244	escherichia
760	4	19.0	30	1	YCCB_ECOLI	Q9mus3	mesostigma
761	4	19.0	33	1	YC12_MESVI	P19937	serratia ma
762	4	19.0	37	1	LPPY_SERMA	P13896	western equ
763	4	19.0	37	1	POLN_WEEV		

#	n	Tm	Len	col	Name	Accession	Organism
764	4	19.0	38	1	CPRP_CANPG	P81033	cancer pagu
765	4	19.0	39	1	PSAG_PEA	P20120	pisum sativ
766	4	19.0	44	1	PHRA_BACSU	Q00829	bacillus su
767	4	19.0	44	1	RK32_LYCES	P36493	lycopersico
768	4	19.0	46	1	HM81_SHEEP	P14159	ovis aries
769	4	19.0	48	1	RL40_METTH	O26653	methanobact
770	4	19.0	48	1	Y234_HAEIN	P43968	haemophilus
771	4	19.0	49	1	LHA7_RHOAC	P35093	rhodopseudo
772	4	19.0	49	1	RK32_ARCFU	O28842	archaeoglob
773	4	19.0	51	1	RK32_ARATH	P42354	arabidopsis
774	4	19.0	54	1	RK32_TOBAC	P12198	nicotiana t
775	4	19.0	56	1	RK32_SPIOL	P28804	spinacia ol
776	4	19.0	57	1	VSH_MUMP2	P28084	mumps virus
777	4	19.0	57	1	VSH_MUMP4	P28085	mumps virus
778	4	19.0	60	1	RCSB_KLEAE	P05339	klebsiella
779	4	19.0	61	1	L503_MOUSE	Q9wvb6	mus musculu
780	4	19.0	61	1	L503_RAT	Q9wvb7	rattus norv
781	4	19.0	62	1	Y4DN_RHISN	P55413	rhizobium s
782	4	19.0	63	1	CE3D_HYPCU	P50720	hyphantria
783	4	19.0	63	1	CE3E_HYPCU	P50721	hyphantria
784	4	19.0	63	1	CE3F_HYPCU	P50722	hyphantria
785	4	19.0	63	1	CE3G_HYPCU	P50723	hyphantria
786	4	19.0	64	1	RL32_RICPR	Q9zch0	rickettsia
787	4	19.0	64	1	VRP1_BPT7	P03704	bacteriopha
788	4	19.0	65	1	AMP1_MELGA	P80391	meleagris g
789	4	19.0	65	1	GLL1_CHICK	P46156	gallus gall
790	4	19.0	65	1	YVC3_CYAPA	P48273	cyanophora
791	4	19.0	65	1	YCE_VACCC	P21123	vaccinia v1
792	4	19.0	66	1	EMX1_MOUSE	Q04742	mus musculu
793	4	19.0	66	1	RPC4_BPP1	P11189	bacteriopha
794	4	19.0	67	1	Y07A_ARCFU	P58023	archaeoglob
795	4	19.0	70	1	ATPL_STRMU	P95783	streptococc
796	4	19.0	70	1	OTX2_RAT	Q64201	thermus aqu
797	4	19.0	71	1	PYRH_THETH	P43891	pachymedusa
798	4	19.0	71	1	DMS2_PACDA	O93454	mus musculu
799	4	19.0	71	1	FBS_MOUSE	Q60791	rickettsia
800	4	19.0	73	1	RL29_RICCN	Q92qx4	rickettsia
801	4	19.0	73	1	RL29_RICPR	Q9zcr3	conus geogr
802	4	19.0	73	1	CXO6_CONGE	P01522	agalychnis
803	4	19.0	73	1	DMS2_AGAAN	O93222	autographa
804	4	19.0	73	1	Y055_NPVAC	P41459	archaeoglob
805	4	19.0	76	1	YG89_ARCFU	O28584	schizophyll
806	4	19.0	77	1	BB1_SCHCO	P78742	rhizobium s
807	4	19.0	78	1	Y4DJ_RHISN	P55409	bacteriopha
808	4	19.0	79	1	Y8K6_BPP22	P14107	streptococc
809	4	19.0	80	1	ATPB_STRDO	P21933	streptococc
810	4	19.0	81	1	R31B_STRPN	Q97qc0	bacillus an
811	4	19.0	82	1	YQ72_BACAN	Q9rmw2	rickettsia
812	4	19.0	83	1	YQ97_RICPR	Q9zcm7	arabidopsis
813	4	19.0	83	1	RS5_ARATH	P51427	mycobacteri
814	4	19.0	83	1	TATA_MYCTU	O10703	escherichia
815	4	19.0	84	1	YQGD_ECOLI	P46879	haliotis la
816	4	19.0	85	1	PLS_HALLA	P82595	vaccinia v1
817	4	19.0	86	1	YVFD_VACCC	P20562	pasteurella
818	4	19.0	86	1	CYS2_PASHA	P72205	streptococc
819	4	19.0	87	1	R31B_STRPY	P9a016	saccharomyc
820	4	19.0	87	1	AGA2_YEAST	P32781	saccharomyc
821	4	19.0	87	1	IM08_YEAST	P57744	pyrococcus
822	4	19.0	87	1	R35A_PYRAB	Q9v1p2	pyrococcus
823	4	19.0	87	1	R35A_PYRHO	O74099	pyrococcus
824	4	19.0	87	1	R35A_PYRWO	Q13106	homo sapien
825	4	19.0	87	1	2154_HUMAN	Q02356	rattus norv
826	4	19.0	88	1	AMD2_RAT	P02371	escherichia
827	4	19.0	88	1	RS15_ECOLI	P44389	haemophilus
828	4	19.0	88	1	RS15_HAEIN	P41120	photorhabdu
829	4	19.0	88	1	RS15_PHOLU	P54079	mycobacteri
830	4	19.0	88	1	TATA_MYCLE	O05113	methylobact
831	4	19.0	88	1	Y113_METEX	P52646	escherichia
832	4	19.0	88	1	YDB1_ECOLI	Q09099	monoraphidi
833	4	19.0	89	1	CYC6_MONBR	P57736	scenedesmus
834	4	19.0	89	1	CYC6_SCEOB	O02414	anthocidari
835	4	19.0	89	1	DYL1_ANTCR	Q22799	caenorhabdi
836	4	19.0	89	1	DYL1_CAEEL		

Left table

No.	SEQ ID				Accession	Organism	Identifier
837	89	1	19.0	4	Q24117	drosophila	DYL1_DROME
838	89	1	19.0	4	Q15701	homo sapien	DYL1_HUMAN
839	89	1	19.0	4	Q96860	drosophila	DYL2_DROME
840	89	1	19.0	4	O67774	aquifex aeo	FLIQ_AQUAE
841	90	1	19.0	4	Q9CK94	pasteurella	DBH_PASMU
842	90	1	19.0	4	P08948	phyllomedus	LITL_PHYSA
843	90	1	19.0	4	P08947	phyllomedus	LITP_PHYSA
844	90	1	19.0	4	O95968	homo sapien	LPPA_HUMAN
845	91	1	19.0	4	Q9S580	chlamydomon	DYL1_CHLRE
846	91	1	19.0	4	P21099	vaccinia vi	VC15_VACCC
847	91	1	19.0	4	P17651	bacteriopha	COR_BPPH8
848	92	1	19.0	4	P70678	mus musculu	HRK_MOUSE
849	92	1	19.0	4	P16161	yersinia pe	Y6D8_YERPE
850	92	1	19.0	4	Q45825	corynebacte	YRI3_CORAM
851	93	1	19.0	4	Q49786	mycobacteri	HIS2_MYCLE
852	93	1	19.0	4	O33257	mycobacteri	HIS2_MYCTU
853	93	1	19.0	4	Q29214	sus scrofa	RLA0_PIG
854	94	1	19.0	4	P27527	petunia hyb	RT19_PETHY
855	95	1	19.0	4	Q9RWB8	deinococcus	MINE_DEIRA
856	95	1	19.0	4	P21050	vaccinia vi	VE10_VACCC
857	95	1	19.0	4	P23373	vaccinia vi	VE10_VACCV
858	95	1	19.0	4	P33821	variola vir	VE10_VARV
859	95	1	19.0	4	Q49760	mycobacteri	Y614_MYCLE
860	96	1	19.0	4	Q9ZJZ9	helicobacte	C553_HELPJ
861	96	1	19.0	4	O25825	helicobacte	C553_HELPY
862	96	1	19.0	4	O37975	bacteriopha	HOLI_BPA18
863	96	1	19.0	4	Q37977	bacteriopha	HOLI_BPA50
864	96	1	19.0	4	Q9G2X0	myxine glut	NULM_MYXGL
865	96	1	19.0	4	O86229	haemophilus	Y87A_HAEIN
866	96	1	19.0	4	P76144	escherichia	YNEC_ECOLI
867	97	1	19.0	4	Q05020	mus musculu	APC2_MOUSE
868	97	1	19.0	4	P03903	mus musculu	NULM_MOUSE
869	97	1	19.0	4	P15554	strongyloce	NULM_STRPU
870	98	1	19.0	4	P00249	nostoc musc	FER2_NOSMU
871	98	1	19.0	4	O79409	scyliorhinu	NULM_SCYCA
872	98	1	19.0	4	Q9ZZ46	squalus aca	NULM_SQUAC
873	98	1	19.0	4	P79342	bos taurus	S113_BOVIN
874	98	1	19.0	4	P97352	mus musculu	S113_MOUSE
875	99	1	19.0	4	P32292	phaseolus a	ARG2_PHAAU
876	99	1	19.0	4	P79209	ovis aries	NOS3_SHEEP
877	99	1	19.0	4	P72759	synechocyst	CCML_SYNY3
878	100	1	19.0	4	Q01709	pseudomonas	NSR2_PSEAE
879	100	1	19.0	4	P05638	pisum sativ	RR14_PEA
880	100	1	19.0	4	Q51507	pseudomonas	PCHB_PSEAE
881	101	1	19.0	4	P81762	pleurotus o	RNPO_PLEOS
882	101	1	19.0	4	P07031	gallus gall	ACYM_CHICK
883	102	1	19.0	4	P49559	odontella s	RK23_ODOSI
884	102	1	19.0	4	Q9BJV9	hadronyche	TXOA_HADIN
885	102	1	19.0	4	P14974	maize strea	Y12K_MSVN
886	103	1	19.0	4	P13204	bos taurus	ANFB_BOVIN
887	103	1	19.0	4	P43455	enterococcu	NTPG_ENTHR
888	103	1	19.0	4	P44362	haemophilus	RL24_HAEIN
889	103	1	19.0	4	Q9C141	pasteurella	RL24_PASMU
890	103	1	19.0	4	P19564	bovine immu	TAT_BIV06
891	103	1	19.0	4	P19565	bovine immu	TAT_BIV27
892	103	1	19.0	4	O32613	bacteriopha	Y01A_BPT4
893	103	1	19.0	4	P22368	claviceps p	YPC6_CLAPU
894	104	1	19.0	4	P47616	mycoplasma	Y376_MYCGE
895	105	1	19.0	4	P06938	human immun	VPR_HV2RO
896	105	1	19.0	4	Q53602	staphylococ	YBXF_STAAU
897	105	1	19.0	4	P71375	haemophilus	YD15_HAEIN
898	106	1	19.0	4	P12485	human immun	REV_HV1BN
899	106	1	19.0	4	P39419	haemophilus	RL11_HAEDU
900	106	1	19.0	4	Q9ZDQ9	rickettsia	Y266_RICPR
901	108	1	19.0	4	P39557	saccharomyc	YAL7_YEAST
902	108	1	19.0	4	P46301	lycopersico	RS25_LYCES
903	109	1	19.0	4	P16123	proteus vul	URE2_PROVU
904	109	1	19.0	4	P79099	bos taurus	AQP2_BOVIN
905	109	1	19.0	4	P79144	canis famil	AQP2_CANFA
906	109	1	19.0	4	P79164	dasypus nov	AQP2_DASNO
907	109	1	19.0	4	O77714	dugong dugo	AQP2_DUGDU
908	109	1	19.0	4	P79165	equus cabal	AQP2_HORSE
909	109	1	19.0	4	P79213	oryctolagus	AQP2_RABIT

Right table

No.	Identifier	SEQ ID				Accession	Organism
910	AQP2_TALEU	109	1	19.0	4	O77740	talpa europ
911	COX3_ARBLI	109	1	19.0	4	Q33752	arbacia lix
912	IATC_CAEEL	109	1	19.0	4	O44441	caenorhabdi
913	YO91_BPP2	109	1	19.0	4	Q06426	bacteriopha
914	AGP1_RAT	110	1	19.0	4	O35460	rattus norv
915	LV1C_MOUSE	110	1	19.0	4	P01725	mus musculu
916	RN_BACCI	110	1	19.0	4	P35078	bacillus ci
917	CO7R_MOUSE	111	1	19.0	4	Q61387	mus musculu
918	LV2F_HUMAN	111	1	19.0	4	P01709	homo sapien
919	LV2H_HUMAN	111	1	19.0	4	P01711	homo sapien
920	Y101_HUMAN	111	1	19.0	4	Q15004	homo sapien
921	Y847_TREPA	111	1	19.0	4	P83819	treponema p
922	YC39_AQUAE	112	1	19.0	4	O67284	aquifex aeo
923	ANFG_AZOMA	112	1	19.0	4	O68946	azomonas ma
924	FLIE_AGRT5	112	1	19.0	4	Q44337	agrobacteri
925	SZ06_BOVIN	112	1	19.0	4	P80221	bos taurus
926	FRT2_HUMAN	113	1	19.0	4	O75474	homo sapien
927	LV1_CHICK	113	1	19.0	4	P04210	gallus gall
928	MEA_PIG	113	1	19.0	4	Q95313	sus scrofa
929	NPFF_HUMAN	113	1	19.0	4	O15130	homo sapien
930	RK20_NEPOL	113	1	19.0	4	Q9TKX4	nephroselmi
931	RSBV_STRCO	113	1	19.0	4	Q9WVX8	streptomyce
932	VL3_BPV4	113	1	19.0	4	P08343	bovine papi
933	YA53_HAEIN	113	1	19.0	4	Q57498	haemophilus
934	YWDK_BACSU	113	1	19.0	4	P39619	bacillus su
935	ALF2_RHOCA	114	1	19.0	4	P50923	rhodobacter
936	KEDA_ACTSL	114	1	19.0	4	P41249	actinomycet
937	MIFH_BRUMA	114	1	19.0	4	P91850	brugia mala
938	MIFH_WUCBA	114	1	19.0	4	O44786	wuchereria
939	RSBV_LISMO	114	1	19.0	4	O85016	listeria mo
940	SZ06_HUMAN	114	1	19.0	4	P80162	homo sapien
941	VG40_BPT4	114	1	19.0	4	P17171	bacteriopha
942	YGI3_BACTU	114	1	19.0	4	P10024	bacillus th
943	COX3_ARTSA	115	1	19.0	4	Q33845	artemia sal
944	IF1A_PYRAB	115	1	19.0	4	Q9V138	pyrococcus
945	Y4HO_RHISN	115	1	19.0	4	P50359	rhizobium s
946	YACC_ECOLI	115	1	19.0	4	P23838	escherichia
947	YIDD_MYCPA	115	1	19.0	4	Q9I7M0	mycobacteri
948	BR31_BRARE	116	1	19.0	4	Q90435	brachydanio
949	BR3A_CHICK	116	1	19.0	4	P55968	gallus gall
950	E411_ADE12	116	1	19.0	4	P36708	human adeno
951	IF1A_PYRHO	116	1	19.0	4	O59280	pyrococcus
952	LV1A_MOUSE	116	1	19.0	4	P01723	mus musculu
953	RL17_SYNP6	117	1	19.0	4	O24711	synechococc
954	SP21_BACME	117	1	19.0	4	P35147	bacillus me
955	SP21_BACST	117	1	19.0	4	O32726	bacillus st
956	ULC9_HCMVA	117	1	19.0	4	P16838	human cytom
957	WN53_EPTST	117	1	19.0	4	P28121	eptatretus
958	EBRB_BACSU	117	1	19.0	4	O31791	bacillus su
959	GHRL_MOUSE	117	1	19.0	4	Q9EQX0	mus musculu
960	GHRL_RAT	117	1	19.0	4	Q9QYH7	rattus norv
961	LVOA_HUMAN	117	1	19.0	4	P04211	homo sapien
962	LV2A_MOUSE	118	1	19.0	4	P01728	mus musculu
963	TAT_SIVS4	118	1	19.0	4	P12513	simian immu
964	WNT6_STRPU	118	1	19.0	4	P28097	strongyloce
965	YDHX_CITFR	118	1	19.0	4	P45516	citrobacter
966	RR19_ASTLO	118	1	19.0	4	P34772	astasia lon
967	YC87_MYCPN	118	1	19.0	4	P75490	mycoplasma
968	IPI_BACSU	119	1	19.0	4	P39804	bacillus su
969	YB82_YEAST	119	1	19.0	4	P38327	saccharomyc
970	YCFF_ECOLI	119	1	19.0	4	P36950	escherichia
971	YJP2_YEAST	119	1	19.0	4	P47003	saccharomyc
972	VG19_BPMU	120	1	19.0	4	Q38646	bacteriopha
973	YCV2_YEAST	120	1	19.0	4	P25363	saccharomyc
974	YJ08_YEAST	120	1	19.0	4	P47106	saccharomyc
975	LSM4_SCHPO	121	1	19.0	4	O14352	schizosacch
976	INI7_HUMAN	122	1	19.0	4	P40305	homo sapien
977	NTF2_ORYSA	122	1	19.0	4	Q9XJ54	oryza sativ
978	PA2N_CROSS	122	1	19.0	4	Q23559	crotalus sc
979	VG20_BPMD2	122	1	19.0	4	Q64214	mycobacteri
980	VG20_BPML5	122	1	19.0	4	Q05226	mycobacteri
981	YMI9_YEAST	122	1	19.0	4	Q04502	saccharomyc
982	CRCB_PYRAB	123	1	19.0	4	Q9V0X2	pyrococcus

```
 983  4  19.0  CRCB_PYRHO  123  1  O59171  pyrococcus
 984  4  19.0  YNL4_YEAST  123  1  P53926  saccharomyc
 985  4  19.0  PAND_AQUAE  124  1  O66773  aquifex aeo
 986  4  19.0  R35A_CAEEL  124  1  P49180  caenorhabdi
 987  4  19.0  RS12_BUCAI  124  1  P57595  buchnera ap
 988  4  19.0  UL39_HCMVA  124  1  P16813  human cytom
 989  4  19.0  VB03_VACCC  124  1  P21000  vaccinia vi
 990  4  19.0  YGDD_HAEIN  124  1  P45019  haemophilus
 991  4  19.0  YQ27_BACAN  124  1  Q9rn05  bacillus an
 992  4  19.0  HV2D_HUMAN  124  1  P01817  homo sapien
 993  4  19.0  RL7_HELPY   125  1  P56875  helicobacte
 994  4  19.0  RL7_HELPJ   125  1  P55834  helicobacte
 995  4  19.0  SSAO_SALTY  125  1  P74858  salmonella
 996  4  19.0  YN25_CAEEL  125  1  P34584  caenorhabdi
 997  4  19.0  CYCP_RHORU  126  1  P00144  rhodospiril
 998  4  19.0  HV2A_HUMAN  126  1  P01814  homo sapien
 999  4  19.0  YIGF_ECOLI  126  1  P27842  escherichia
1000  4  19.0  YIGF_SALTY  126  1  P31139  salmonella
```

ALIGNMENTS

```
RESULT 1
FIL1_YEAST
ID   FIL1_YEAST     STANDARD;      PRT;     230 AA.
AC   P3877l;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   FIL1 protein, mitochondrial precursor.
GN   FIL1 OR KIM4 OR YHR038W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT8/1;
RX   MEDLINE=98417448; PubMed=9746366;
RA   Kanai T., Takeshita S., Atomi H., Umemura K., Ueda M., Tanaka A.;
RT   "A regulatory factor, Fil1p, involved in derepression of the
RT   isocitrate lyase gene in Saccharomyces cerevisiae -- a possible
RT   mitochondrial protein necessary for protein synthesis in
RT   mitochondria.";
RL   Eur. J. Biochem. 256:212-220(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=9478003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latrelle P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [3]
RP   SIMILARITY TO RRF.
RX   MEDLINE=96155066; PubMed=8563640;
RA   Ouzounis C., Bork P., Casari G., Sander C.;
RT   "New protein functions in yeast chromosome VIII.";
RL   Protein Sci. 4:2424-2428(1995).
CC   -!- FUNCTION: INVOLVED IN GLUCOSE REPRESSION/DEREPRESSION. MAY BE
CC       NECESSARY FOR PROTEIN SYNTHESIS IN MITOCHONDRIA. MAY FUNCTION AS A
CC       RIBOSOME RECYCLING FACTOR IN MITOCHONDRIA.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE RRF FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AB016033; BAA31687.1; -.
DR   EMBL; U00062; AAB68906.1; -.
DR   PIR; S46737; S46737.
DR   SGD; S0001080; FIL1.
DR   InterPro; IPR002661; RRF.
DR   Pfam; PF01765; RRF; 1.
KW   Protein biosynthesis; Transit peptide; Mitochondrion.
FT   TRANSIT       1     24       MITOCHONDRION (POTENTIAL).
FT   CHAIN        25    230       FIL1 PROTEIN.
SQ   SEQUENCE   230 AA;  26406 MW;  9CBCD8C5FB6F3008 CRC64;

Query Match             33.3%;  Score 7;  DB 1;  Length 230;
Best Local Similarity   100.0%;  Pred. No. 2;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps     0;

Qy      9 RLFNRSF 15
          |||||||
Db     17 RLFNRSF 23


RESULT 2
ABF1_TRIRE
ID   ABF1_TRIRE     STANDARD;      PRT;     500 AA.
AC   Q92455;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN   ABF1.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=QM9414 / RUT C-30;
RX   MEDLINE=96434473; PubMed=8837440;
RA   Margolles-Clark E., Tenkanen M., Nakari-Setaelae T., Penttilae M.;
RT   "Cloning of genes encoding alpha-L-arabinofuranosidase and beta-
RT   xylosidase from Trichoderma reesei by expression in Saccharomyces
RT   cerevisiae.";
RL   Appl. Environ. Microbiol. 62:3840-3846(1996).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC       arabinofuranoside residues in alpha-L-arabinosides.
CC   -!- PATHWAY: INVOLVED IN DEGRADATION OF THE PLANT CELL WALL
CC       POLYSACCHARIDE L-ARABINAN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z69252; CAA93243.1; -.
KW   Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    500       ALPHA-L-ARABINOFURANOSIDASE.
FT   CARBOHYD    467    467       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   500 AA;  51115 MW;  FC6B4DB03EE3C762 CRC64;

Query Match             33.3%;  Score 7;  DB 1;  Length 500;
Best Local Similarity   100.0%;  Pred. No. 4;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps     0;
```

Qy   1 SGSLSTF 7
         |||||||
Db   266 SGSLSTF 272

RESULT 3
XYL1_TRIKO
ID   XYL1_TRIKO   STANDARD;   PRT;   500 AA.
AC   P48792;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Arabinofuranosidase/B-xylosidase precursor (Includes: Alpha-L-
DE   arabinofuranosidase (EC 3.2.1.55) (Arabinosidase); Beta-xylosidase
DE   (EC 3.2.1.37) (1,4-beta-D-xylan xylohydrolase) (Xylan 1,4-beta-
DE   xylosidase)).
GN   XYL1.
OS   Trichoderma koningii.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=55202;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G-39;
RA   Huang L.N., Hseu T.H., Lee Y.J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC   arabinofuranoside residues in alpha-L-arabinosides.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-XYLANS SO AS TO
CC   REMOVE SUCCESSIVE D-XYLOSE RESIDUES FROM THE NON-REDUCING
CC   TERMINI. IT ALSO HYDROLYSES XYLOBIOSE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 54 OF GLYCOSYL HYDROLASES.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U38661; AAA81024.1; -.
KW   Xylan degradation; Hydrolase; Glycosidase; Multifunctional enzyme;
KW   Glycoprotein; Signal.
FT   SIGNAL   1   21   POTENTIAL.
FT   CHAIN   22   500   ARABINOFURANOSIDASE/B-XYLOSIDASE.
FT   CARBOHYD   467   467   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   500 AA;   51129 MW; 06DFC319AAFA1149 CRC64;

Query Match   33.3%; Score 7; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 4;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLSTF 7
         |||||||
Db   266 SGSLSTF 272

RESULT 4
Y805_AQUAE
ID   Y805_AQUAE   STANDARD;   PRT;   155 AA.
AC   O66989;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_805.
GN   AQ_805.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE000708; AAC06954.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   155 AA;   18152 MW; 16933A2AC734531 CRC64;

Query Match   28.6%; Score 6; DB 1; Length 155;
Best Local Similarity 100.0%; Pred. No. 15;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   10 LFNRSF 15
         ||||||
Db   54 LFNRSF 59

RESULT 5
CCAC_CAVPO
ID   CCAC_CAVPO   STANDARD;   PRT;   169 AA.
AC   O35505;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Voltage-dependent L-type calcium channel alpha-1C subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 1, cardiac muscle)
DE   (Fragment).
GN   CACNA1C OR CACNL1A1 OR CCHL1A1 OR CACH2 OR CACN2.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Collins P.L., Lundgren D.W., Kulp T.M., Shah P., Chang S.M.,
RA   Chang A.S.;
RT   "Gestational expression of voltage-dependent calcium channel subunits
RT   in guinea pig uterus.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC   ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC   IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC   CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC   CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1C
CC   GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)
CC   CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)
CC   GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP),
CC   PHENYLALKYLAMINES, BENZOTHIAZEPINES, AND BY OMEGA-AGATOXIN-IIIA
CC   (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-
CC   GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
CC   CALCIUM CHANNELS CONTAINING THE ALPHA-1C SUBUNIT PLAY AN IMPORTANT
CC   ROLE IN EXCITATION-CONTRACTION COUPLING IN THE HEART (BY
CC   SIMILARITY).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC   COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC   IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC   FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC   SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC   CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Basilar papilla;
RX   MEDLINE=98070847; PubMed=9405708;
RA   Kollmar R., Montgomery L.G., Fak J., Henry L.J., Hudspeth A.J.;
RT   "Predominance of the alpha1D subunit in L-type voltage-gated Ca2+
RT   channels of hair cells in the chicken's cochlea.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14883-14888(1997).
CC   -!- FUNCTION: THE ISOFORM ALPHA-1C GIVES RISE TO L-TYPE CALCIUM
CC       CURRENTS. LONG-LASTING (L-TYPE) CALCIUM CHANNELS BELONG TO THE
CC       "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP. CALCIUM CHANNELS CONTAINING
CC       THE ALPHA-1C SUBUNIT PLAY AN IMPORTANT ROLE IN EXCITATION-
CC       CONTRACTION COUPLING IN THE HEART (BY SIMILARITY).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- DOMAIN: BINDING OF INTRACELLULAR CALCIUM THROUGH THE EF-HAND MOTIF
CC       INHIBITS THE OPENING OF THE CHANNEL (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATION BY CAPK ACTIVATES THE CHANNEL (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF005938; AAB62890.1; -.
DR   EMBL; AF027610; AAC08311.1; -.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation.
FT   NON_TER        1      1
FT   REPEAT        <1    >177       IV.
FT   TRANSMEM      <1     13        S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        14     26        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      27     45        S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        46     64        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      65     84        S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        85    153        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     154   >177        S6 OF REPEAT IV (POTENTIAL).
FT   BINDING      132   >177        TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      146   >177        TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      118   >177        TO PHENYLALKYLAMINES (BY SIMILARITY).
FT   SITE         118             CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                               (BY SIMILARITY).
FT   CARBOHYD      90     90        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     141    141        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER      177    177
SQ   SEQUENCE   177 AA;   19957 MW;   84CB4656D78AAF23 CRC64;

Query Match             28.6%;  Score 6; DB 1;  Length 177;
Best Local Similarity  100.0%;  Pred. No. 17;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- DOMAIN: BINDING OF INTRACELLULAR CALCIUM THROUGH THE EF-HAND MOTIF
CC       INHIBITS THE OPENING OF THE CHANNEL (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATION BY CAPK ACTIVATES THE CHANNEL (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF005938; AAB62890.1; -.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation.
FT   NON_TER        1      1
FT   TRANSMEM      <1     11        S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        12     38        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      39     57        S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        58     76        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      77     96        S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        97    165        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     166   >169        S6 OF REPEAT IV (POTENTIAL).
FT   BINDING       <1     11        TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      144   >169        TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      158   >169        TO PHENYLALKYLAMINES (BY SIMILARITY).
FT   SITE         130             CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                               (BY SIMILARITY).
FT   CARBOHYD     102    102        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     153    153        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER      169    169
SQ   SEQUENCE   169 AA;   19514 MW;   13E88E8510D30C88 CRC64;

Query Match             28.6%;  Score 6; DB 1;  Length 169;
Best Local Similarity  100.0%;  Pred. No. 17;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     6 TFFRLF 11
         ||||||
Db    40 TFFRLF 45

RESULT 6
CCAC_CHICK
ID   CCAC_CHICK     STANDARD;     PRT;    177 AA.
AC   O73707;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Voltage-dependent L-type calcium channel alpha-1C subunit (CHCACHA1C)
DE   (Fragment).
GN   CACNA1C.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
```

```
Qy    6 TFFRLF 11
        ||||||
Db   28 TFFRLF 33

RESULT 7
ID   Y540_AQUAE      STANDARD;      PRT;   214 AA.
AC   O66819;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_540.
GN   AQ_540.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0056 (MARC) FAMILY.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; AE000694; AAC06780.1; -.
DR   InterPro; IPR002771; UPF0056.
DR   Pfam; PF01914; UPF0056; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     17     37       POTENTIAL.
FT   TRANSMEM     47     67       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    153    173       POTENTIAL.
FT   TRANSMEM    185    205       POTENTIAL.
SQ   SEQUENCE   214 AA;  23572 MW;  DD1A9947A5DE931D CRC64;

Query Match     28.6%;  Score 6;  DB 1;  Length 214;
Best Local Similarity  100.0%;  Pred. No. 20;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    3 SLSTFF 8
        ||||||
Db  171 SLSTFF 176

RESULT 8
ID   CCAD_MOUSE      STANDARD;      PRT;   281 AA.
AC   Q99246;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Voltage-dependent L-type calcium channel alpha-1D subunit (Calcium
DE   channel, L type, alpha-1 polypeptide isoform 2) (Fragment).
GN   CACNA1D OR CACNL1A2 OR CCHL1A2 OR CACH3 OR CACN4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM CACH3B).
RC   STRAIN=ICR; TISSUE=Ovary;
RX   MEDLINE=91056091; PubMed=2173707;
RA   Perez-Reyes E., Wei X., Castellano A., Birnbaumer L.;
RT   "Molecular diversity of L-type calcium channels. Evidence for
RT   alternative splicing of the transcripts of three non-allelic genes.";
RL   J. Biol. Chem. 265:20430-20436(1990).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1D
CC       GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)
CC       CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)
CC       GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP),
CC       PHENYLALKYLAMINES, BENZOTHIAZEPINES, AND BY OMEGA-AGATOXIN-IIIA
CC       (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-
CC       GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: MULTIPLE ISOFORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING (PROBABLE). THE SEQUENCE SHOWN HERE IS THAT
CC       OF CACH3B.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; M57975; AAA63292.1; -.
DR   MGD; MGI:88293; Cacna1d.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Alternative splicing.
FT   NON_TER       1      1
FT   REPEAT       45    >281       IV.
FT   DOMAIN        2     58       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     59     77       S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN       78     92       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     93    112       S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN      113    119       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    120    141       S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN      142    151       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    152    171       S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN      172    190       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    191    210       S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN      211    277       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    278   >281       S6 OF REPEAT IV (POTENTIAL).
FT   SITE        244    244       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   BINDING      <1     18       TO DIHYDROPYRIDINES (BY SIMILARITY).
```

```
FT   BINDING       258    >281       TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING       270    >281       TO PHENYLALKYLAMINES (BY SIMILARITY).
FT   NON_TER       281    281
SQ   SEQUENCE      281 AA;  32615 MW;  6F317F54D81E3F6F CRC64;

Query Match             28.6%;  Score 6;  DB 1;  Length 281;
Best Local Similarity   100.0%;  Pred. No. 26;
Matches      6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6  TFFRLF  11
           ||||||
Db    154  TFFRLF  159

RESULT 9
CCAS_CHICK
ID   CCAS_CHICK   STANDARD;       PRT;    281 AA.
AC   O42398;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Voltage-dependent L-type calcium channel alpha-1S subunit (Fragment).
GN   CACNA1S.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Intestinal epithelium;
RA   Xu J., Norman A.W., Henry H.L., de Boland A.R., Zanello L.P.;
RT   "Molecular characterization of an L-type calcium channel in chick
RT   intestinal epithelia cells.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE ISOFORM ALPHA-1S GIVES RISE TO L-TYPE CALCIUM
CC       CURRENTS. LONG-LASTING (L-TYPE) CALCIUM CHANNELS BELONG TO THE
CC       "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP. THEY ARE BLOCKED BY
CC       DIHYDROPYRIDINES (DHP), PHENYLALKYLAMINES, BENZOTHIAZEPINES, AND
CC       BY OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA). THEY ARE HOWEVER
CC       INSENSITIVE TO OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND OMEGA-
CC       AGATOXIN-IVA (OMEGA-AGA-IVA). CALCIUM CHANNELS CONTAINING THE
CC       ALPHA-1S SUBUNIT PLAY AN IMPORTANT ROLE IN EXCITATION- CONTRACTION
CC       COUPLING IN SKELETAL MUSCLE (BY SIMILARITY).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY. AN
CC       ADDITIONAL GAMMA SUBUNIT IS PRESENT ONLY IN SKELETAL MUSCLE L-TYPE
CC       CHANNEL (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- DOMAIN: THE LOOP BETWEEN REPEATS II AND III INTERACTS WITH THE
CC       RYANODINE RECEPTOR, AND IS THEREFORE IMPORTANT FOR CALCIUM RELEASE
CC       FROM THE ENDOPLASMIC RETICULUM NECESSARY FOR MUSCLE CONTRACTION
CC       (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATION BY CAPK STIMULATES THE CALCIUM CHANNEL
CC       FUNCTION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------
```

```
DR   EMBL; AF007877; AAB63206.1; -.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation.
FT   NON_TER        1      1
FT   REPEAT        <1      8        III.
FT   REPEAT        45    >281       IV.
FT   TRANSMEM      <1      5        S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN         6     58        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      59     77        S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN        78     92        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      93    112        S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN       113    120        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     121    139        S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN       140    152        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     153    171        S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN       172    190        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     191    210        S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN       211    277        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     278    >281       S6 OF REPEAT IV (POTENTIAL).
FT   BINDING       <1     17        TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      258    >281       TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING      270    >281       TO PHENYLALKYLAMINES (BY SIMILARITY).
FT   SITE         244    244        CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                  (BY SIMILARITY).
FT   CARBOHYD      81     81        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER      281    281
SQ   SEQUENCE     281 AA;  32729 MW;  46BE70FA44D9166B CRC64;

Query Match             28.6%;  Score 6;  DB 1;  Length 281;
Best Local Similarity   100.0%;  Pred. No. 26;
Matches      6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6  TFFRLF  11
           ||||||
Db    154  TFFRLF  159

RESULT 10
SCF_CHICK
ID   SCF_CHICK    STANDARD;       PRT;    287 AA.
AC   Q09108;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kit ligand precursor (C-kit ligand) (Stem cell factor) (SCF) (Mast
DE   cell growth factor) (MGF).
GN   KITLG.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93273244; PubMed=7684722;
RA   Zhou J., Ohtaki M., Sakurai M.;
RT   "Sequence of a cDNA encoding chicken stem cell factor.";
RL   Gene 127:269-270(1993).
CC   -!- FUNCTION: STIMULATES THE PROLIFERATION OF MAST CELLS. ABLE TO
CC       AUGMENT THE PROLIFERATION OF BOTH MYELOID AND LYMPHOID
CC       HEMATOPOIETIC PROGENITORS IN BONE MARROW CULTURE. MEDIATES ALSO
CC       CELL-CELL ADHESION. ACTS SYNERGISTICALLY WITH OTHER CYTOKINES,
CC       PROBABLY INTERLEUKINS.
```

```
CC    -!- SUBUNIT: HOMODIMER, NON-COVALENTLY LINKED (PROBABLE).
CC    -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO EXISTS AS A
CC        SECRETED SOLUBLE FORM.
CC    -!- DEVELOPMENTAL STAGE: ACTS IN THE EARLY STAGES OF HEMATOPOIESIS.
CC    -!- SIMILARITY: BELONGS TO THE SCF FAMILY.
CC    -------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -------------------------------------------------------
DR    EMBL; D13516; BAA02733.1; -.
DR    PIR; JN0637; JN0637.
DR    InterPro; IPR003452; SCF.
DR    Pfam; PF02404; SCF; 1.
KW    Growth factor; Glycoprotein; Transmembrane; Signal; Cell adhesion.
FT    SIGNAL       1     25       POTENTIAL.
FT    CHAIN       26    287       KIT LIGAND.
FT    DOMAIN      26    225       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM   226    246       POTENTIAL.
FT    DOMAIN     247    287       CYTOPLASMIC (POTENTIAL).
FT    DISULFID    29    117       BY SIMILARITY.
FT    DISULFID    68    167       BY SIMILARITY.
FT    CARBOHYD   100    100       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD   149    149       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD   200    200       N-LINKED (GLCNAC...) (POTENTIAL).
SQ    SEQUENCE   287 AA;  32328 MW;  6AEB556ADC152578 CRC64;

Query Match             28.6%;   Score 6;   DB 1;   Length 287;
Best Local Similarity   100.0%;  Pred. No. 27;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      7 FFRLFN 12
          ||||||
Db    144 FFRLFN 149

RESULT 11
CEBB_MOUSE
ID    CEBB_MOUSE     STANDARD;     PRT;    296 AA.
AC    P28033;
DT    01-AUG-1992 (Rel. 23, Created)
DT    01-AUG-1992 (Rel. 23, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    CCAAT/enhancer binding protein beta (C/EBP beta) (Interleukin-6-
DE    dependent binding protein) (IL-6DBP) (Liver-enriched transcriptional
DE    activator) (LAP) (AGP/EBP).
GN    CEBPB.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=BALB/C; TISSUE=Liver;
RX    MEDLINE=91061773; PubMed=1701020;
RA    Chang C.J., Chen T.T., Lei H.Y., Chen D.S., Lee S.C.;
RT    "Molecular cloning of a transcription factor, AGP/EBP, that belongs
RT    to members of the C/EBP family.";
RL    Mol. Cell. Biol. 10:6642-6653(1990).
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=91357470; PubMed=1840554;
RA    Cao Z., Umek R.M., McKnight S.L.;
RT    "Regulated expression of three C/EBP isoforms during adipose
RT    conversion of 3T3-L1 cells.";
RL    Genes Dev. 5:1538-1552(1991).
RN    [3]
RP    SEQUENCE OF 1-11 FROM N.A.
RC    TISSUE=Liver;
```

```
RX    MEDLINE=95322031; PubMed=7598808;
RA    Chang C.J., Shen B.J., Lee S.C.;
RT    "Autoregulated induction of the acute-phase response transcription
RT    factor gene, agp/ebp.";
RL    DNA Cell Biol. 14:529-537(1995).
CC    -!- FUNCTION: INVOLVED IN INTERLEUKIN-6 SIGNAL TRANSDUCTION,
CC        INCLUDING THE TRANSCRIPTIONAL ACTIVATION OF ACUTE-PHASE GENES.
CC    -!- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC        WITH C/EBP ALPHA AND GAMMA.
CC    -!- SUBCELLULAR LOCATION: Nuclear.
CC    -!- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC        PROTEINS.
CC    -------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -------------------------------------------------------
DR    EMBL; X62600; CAA44484.1; -.
DR    EMBL; M61007; AAA37192.1; -.
DR    EMBL; S78572; -; NOT_ANNOTATED_CDS.
DR    PIR; A37279; A37279.
DR    PIR; A36366; A36366.
DR    TRANSFAC; T00017; -.
DR    MGD; MGI:88373; Cebpb.
DR    InterPro; IPR001871; bZIP.
DR    Pfam; PF00170; bZIP; 1.
DR    SMART; SM00338; BRLZ; 1.
KW    Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW    Trans-acting factor.
FT    DOMAIN     120    129       PRO-RICH.
FT    DOMAIN     170    191       PRO/SER-RICH.
FT    DNA_BIND   228    244       BASIC MOTIF.
FT    DOMAIN     257    285       LEUCINE-ZIPPER.
SQ    SEQUENCE   296 AA;  31445 MW;  827AC4AFC209AEB9 CRC64;

Query Match             28.6%;   Score 6;   DB 1;   Length 296;
Best Local Similarity   100.0%;  Pred. No. 27;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 SGSLST 6
          ||||||
Db    174 SGSLST 179

RESULT 12
CEBB_RAT
ID    CEBB_RAT       STANDARD;     PRT;    297 AA.
AC    P21272;
DT    01-MAY-1991 (Rel. 18, Created)
DT    01-MAY-1991 (Rel. 18, Last sequence update)
DT    15-JUL-1998 (Rel. 36, Last annotation update)
DE    CCAAT/enhancer binding protein beta (C/EBP beta) (Interleukin-6-
DE    dependent binding protein) (IL-6DBP) (Liver-enriched transcriptional
DE    activator) (LAP) (Silencer factor B) (SF-B) (C/EBP-related protein 2).
GN    CEBPB OR SFB OR CRP2.
OS    Rattus norvegicus (Rat).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX    NCBI_TaxID=10116;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=91029495; PubMed=2171780;
RA    Poli V., Mancini F.P., Cortese R.;
RT    "IL-6DBP, a nuclear protein involved in interleukin-6 signal
RT    transduction, defines a new family of leucine zipper proteins related
RT    to C/EBP.";
RL    Cell 63:643-653(1990).
RN    [2]
RP    SEQUENCE FROM N.A.
```

RESULT 13
ISPB_HAEIN
ID   ISPB_HAEIN   STANDARD;   PRT;   329 AA.
AC   P44916;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Octaprenyl-diphosphate synthase (EC 2.5.1.-) (Octaprenyl pyrophosphate
DE   synthetase) (OPP synthetase).
GN   ISPB OR HI0881.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: SUPPLIES OCTAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE
CC   SIDE CHAIN OF THE ISOPRENOID QUINONES UBIQUINONE AND MENAQUINONE
CC   (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U32770; AAC22540.1; -.
DR   TIGR; HI0881; -.
DR   InterPro; IPR000092; Polyprenyl_synt.
DR   Pfam; PF00348; polyprenyl_synt; 1.
DR   PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR   PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
KW   Isoprene biosynthesis; Transferase; Complete proteome.
SQ   SEQUENCE 329 AA; 35911 MW; 65DBC4CCDAD72E04 CRC64;

Query Match   28.6%; Score 6; DB 1; Length 329;
Best Local Similarity   100.0%; Pred. No. 30;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   16 TQALGK 21
            ||||||
Db   226 TQALGK 231

RESULT 14
CEBB_HUMAN
ID   CEBB_HUMAN   STANDARD;   PRT;   345 AA.
AC   P17676; Q9H4Z5;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   CCAAT/enhancer binding protein beta (C/EBP beta) (Nuclear factor
DE   NF-IL6) (Transcription factor 5).
GN   CEBPB OR TCF5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

RC   STRAIN=LEWIS; TISSUE=Liver;
RX   MEDLINE=91071582; PubMed=2253878;
RA   Descombes P., Chojkier M., Lichtsteiner S., Falvey E., Schibler U.;
RT   "LAP, a novel member of the C/EBP gene family, encodes a
RT   liver-enriched transcriptional activator protein.";
RL   Genes Dev. 4:1541-1551(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=92319638; PubMed=1377818;
RA   Thomassin H., Hamel D., Bernier D., Guertin M., Belanger L.;
RT   "Molecular cloning of two C/EBP-related proteins that bind to the
RT   promoter and the enhancer of the alpha 1-fetoprotein gene. Further
RT   analysis of C/EBP beta and C/EBP gamma.";
RL   Nucleic Acids Res. 20:3091-3098(1992).
RN   [4]
RP   SEQUENCE OF 77-297 FROM N.A.
RC   TISSUE=Liver;
RA   Imagawa M., Osada S., Koyama Y., Suzuki T., Hirom P.C.,
RA   Diccianni M.B., Morimura S., Muramatsu M.;
RL   Submitted (JUL-1991) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 22-297 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver, Adipose tissue, and Lung;
RX   MEDLINE=91357747; PubMed=1884998;
RA   Williams S.C., Cantwell C.A., Johnson P.F.;
RT   "A family of C/EBP-related proteins capable of forming covalently
RT   linked leucine zipper dimers in vitro.";
RL   Genes Dev. 5:1553-1567(1991).
CC   -!- FUNCTION: INVOLVED IN INTERLEUKIN-6 SIGNAL TRANSDUCTION,
CC   INCLUDING THE TRANSCRIPTIONAL ACTIVATION OF ACUTE-PHASE GENES.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC   WITH C/EBP ALPHA AND GAMMA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: LIVER AND LUNG.
CC   -!- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC   PROTEINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M57235; AAA19669.1; -.
DR   EMBL; X54626; CAA38443.1; -.
DR   EMBL; X60769; CAA43179.1; -.
DR   EMBL; M84011; AAA40972.1; -.
PIR; A36197; A36197.
PIR; A35914; A35914.
TRANSFAC; T00459; -.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRL2; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Trans-acting factor.
FT   DOMAIN   121   130   PRO-RICH.
FT   DOMAIN   171   192   PRO/SER-RICH.
FT   DNA_BIND   229   245   BASIC MOTIF.
FT   DOMAIN   258   286   LEUCINE-ZIPPER.
SQ   SEQUENCE 297 AA; 31502 MW; C2511FDB65527789 CRC64;

Query Match   28.6%; Score 6; DB 1; Length 297;
Best Local Similarity   100.0%; Pred. No. 27;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 SGSLST 6
            ||||||
Db   175 SGSLST 180

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90269225; PubMed=2112087;
RA   Akira S., Isshiki H., Sugita T., Tanabe O., Kinoshita S.,
RA   Nishio Y., Nakajima T., Hirano T., Kishimoto T.;
RT   "A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP
RT   family.";
RL   EMBO J. 9:1897-1906(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaslaiho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C.A., Steward C.A., Sulston J.E.,
RA   Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
CC   -!- FUNCTION: IMPORTANT TRANSCRIPTIONAL ACTIVATOR IN THE REGULATION OF
CC       GENES INVOLVED IN IMMUNE AND INFLAMMATORY RESPONSES. SPECIFICALLY
CC       BINDS TO AN IL-1 RESPONSE ELEMENT IN THE IL-6 GENE. NF-IL6 ALSO
CC       BINDS TO REGULATORY REGIONS OF SEVERAL ACUTE-PHASE AND CYTOKINES
CC       GENES. IT PROBABLY PLAYS A ROLE IN THE REGULATION OF ACUTE-PHASE
CC       REACTION, INFLAMMATION AND HEMOPOIESIS. THE CONSENSUS RECOGNITION
CC       SITE IS 5'-T(T/G)NNGNAA(T/G)-3'.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC       WITH C/EBP ALPHA, DELTA AND GAMMA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN THE LUNG, KIDNEY,
CC       AND SPLEEN.
CC   -!- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC       PROTEINS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X52560; CAA36794.1; -.
DR   EMBL; AL161937; CAC14276.1; -.
DR   PIR; S12788; S12788.
DR   TRANSFAC; T00581; -.
DR   MIM; 189965; -.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DOMAIN       162    170       POLY-PRO.
FT   DNA_BIND     277    293       BASIC MOTIF.
FT   DOMAIN       306    334       LEUCINE-ZIPPER.
FT   CONFLICT     253    253       G -> A (IN REF. 2).
SQ   SEQUENCE   345 AA;  36091 MW;  C4D7A476CACC717D CRC64;

Query Match             28.6%;  Score 6;  DB 1;  Length 345;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SGSLST 6
          ||||||
Db    221 SGSLST 226

RESULT 15
CEBB_BOVIN
ID   CEBB_BOVIN     STANDARD;      PRT;     348 AA.
AC   O02755;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   CCAAT/enhancer binding protein beta (C/EBP beta).
GN   CEBPB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Japanese black;
RX   MEDLINE=97203913; PubMed=9051485;
RA   Yamaoka I., Taniguchi Y., Sasaki Y.;
RT   "Nucleotide sequence of bovine C/EBP beta gene.";
RL   J. Anim. Sci. 75:587-587(1997).
CC   -!- FUNCTION: INVOLVED IN INTERLEUKIN-6 SIGNAL TRANSDUCTION, INCLUDING
CC       THE TRANSCRIPTIONAL ACTIVATION OF ACUTE-PHASE GENES (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC       WITH C/EBP ALPHA AND GAMMA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC       PROTEINS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; D82985; BAA20096.1; -.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DOMAIN        13     19       POLY-PRO.
FT   DOMAIN       162    170       POLY-PRO.
FT   DOMAIN       224    232       POLY-SER.
FT   DOMAIN       247    250       POLY-ALA.
FT   DNA_BIND     280    296       BASIC MOTIF.
FT   DOMAIN       309    337       LEUCINE-ZIPPER.
SQ   SEQUENCE   348 AA;  36390 MW;  7059D2F87B474CB4 CRC64;

Query Match             28.6%;  Score 6;  DB 1;  Length 348;
Best Local Similarity  100.0%;  Pred. No. 32;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SGSLST 6
          ||||||
Db    222 SGSLST 227
```

us-09-833-017b-4.oligo.rsp

Wed Nov 13 14:19:10 2002

Search completed: November 5, 2002, 11:05:40
Job time : 20.6418 secs

Copyright   GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2002, 10:58:27 ; Search time 13.791 Seconds
(without alignments)
263.424 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 21
Sequence: 1 SGSLSTFFRLFNRSFTQALGK 21
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172294929 residues
Word size :   0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :   SPTREMBL_19:*
  1: sp_archea:*
  2: sp_bacteria:*
  3: sp_fungi:*
  4: sp_human:*
  5: sp_invertebrate:*
  6: sp_mammal:*
  7: sp_mhc:*
  8: sp_organelle:*
  9: sp_phage:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_virus:*
 13: sp_vertebrate:*
 14: sp_unclassified:*
 15: sp_rvirus:*
 16: sp_bacteriap:*
 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	46	2	Q99Q15	Q99q15 streptococc
2	21	100.0	46	2	Q9APK7	Q9apk7 streptococc
3	15	71.4	43	2	Q9APK6	Q9apk6 streptococc
4	7	33.3	253	13	Q90315	Q90315 coturnix co
5	7	33.3	287	13	Q90314	Q90314 coturnix co
6	7	33.3	929	5	Q9VR32	Q9vr32 drosophila
7	6	28.6	51	12	Q91G53	Q91g53 chilo iride
8	6	28.6	69	12	O41074	O41074 paramecium
9	6	28.6	69	12	Q91PP4	Q91pp4 hepatitis b
10	6	28.6	122	16	Q9RU75	Q9ru75 deinococcus
11	6	28.6	134	12	Q84484	Q84484 paramecium
12	6	28.6	140	4	Q9BSC0	Q9bsc0 homo sapien
13	6	28.6	177	15	Q9QRM3	Q9qrm3 human immun
14	6	28.6	180	16	Q98FX0	Q98fx0 rhizobium l
15	6	28.6	181	16	Q901F3	Q901f3 human immun
16	6	28.6	186	10	O04571	O04571 arabidopsis
17	6	28.6	186	11	Q62146	Q62146 mus musculu
18	6	28.6	191	11	Q9DOT7	Q9DOt7 mus musculu
19	6	28.6	200	6	O16314	O16314 caenorhabdi
20	6	28.6	200	6	Q95LH6	Q95LH6 bos taurus
21	6	28.6	201	15	Q83184	Q83184 mink cell f
22	6	28.6	202	16	Q9PLC4	Q9PLC4 chlamydia m
23	6	28.6	204	11	Q9EPW1	Q9EPW1 rattus norv
24	6	28.6	205	5	O15772	O15772 trypanosoma
25	6	28.6	205	5	O15777	O15777 trypanosoma
26	6	28.6	213	16	O84471	O84471 chlamydia t
27	6	28.6	246	2	Q18894	Q18894 caenorhabdi
28	6	28.6	249	2	Q9RB46	Q9RB46 clostridium
29	6	28.6	271	16	Q9K8F9	Q9K8F9 bacillus ha
30	6	28.6	285	16	Q9KDP6	Q9KDP6 bacillus ha
31	6	28.6	287	16	Q97FG4	Q97FG4 clostridium
32	6	28.6	288	16	Q9PB74	Q9PB74 xylella fas
33	6	28.6	289	5	O01509	O01509 caenorhabdi
34	6	28.6	297	11	Q64106	Q64106 rattus sp.
35	6	28.6	300	16	Q93UN6	Q93UN6 vibrio vuln
36	6	28.6	300	16	Q91259	Q91259 pseudomonas
37	6	28.6	314	1	P94910	P94910 methanosarc
38	6	28.6	318	4	Q9UMU8	Q9UMU8 homo sapien
39	6	28.6	329	16	O83509	O83509 treponema p
40	6	28.6	336	16	Q87519	Q87519 escherichia
41	6	28.6	340	3	P78846	P78846 schizosacch
42	6	28.6	342	3	Q9GRL0	Q9GRL0 leishmania
43	6	28.6	345	4	Q96IH2	Q96IH2 homo sapien
44	6	28.6	359	5	Q9V7Q2	Q9V7Q2 drosophila
45	6	28.6	360	16	Q9Z8L6	Q9Z8L6 chlamydia p
46	6	28.6	370	10	Q9LYK0	Q9LYK0 arabidopsis
47	6	28.6	421	16	Q92WM5	Q92WM5 rhizobium m
48	6	28.6	425	10	Q9MA53	Q9MA53 arabidopsis
49	6	28.6	465	10	Q9FHT1	Q9FHT1 arabidopsis
50	6	28.6	466	5	Q15846	Q15846 homo sapien
51	6	28.6	472	3	Q9U082	Q9U082 schizosacch
52	6	28.6	478	5	Q9WOH5	Q9WOH5 drosophila
53	6	28.6	484	5	Q26475	Q26475 schistocerc
54	6	28.6	496	2	Q9AD76	Q9AD76 streptomyce
55	6	28.6	499	3	Q9C4B1	Q9C4B1 aspergillus
56	6	28.6	499	3	Q12551	Q12551 aspergillus
57	6	28.6	506	3	Q96X02	Q96X02 penicillium
58	6	28.6	506	3	Q96VA1	Q96VA1 aspergillus
59	6	28.6	507	3	Q96VA0	Q96VA0 aspergillus
60	6	28.6	510	4	Q96CL7	Q96CL7 homo sapien
61	6	28.6	511	4	O74288	O74288 emericella
62	6	28.6	519	10	Q942V7	Q942V7 oryza sativ
63	6	28.6	534	5	Q22017	Q22017 caenorhabdi
64	6	28.6	540	5	Q961B3	Q961B3 drosophila
65	6	28.6	559	2	O68145	O68145 burkholderl
66	6	28.6	561	2	Q9NVH1	Q9NVH1 homo sapien
67	6	28.6	575	2	Q9EYF3	Q9EYF3 escherichia
68	6	28.6	583	5	O45066	O45066 caenorhabdi
69	6	28.6	583	5	O61711	O61711 caenorhabdi
70	6	28.6	590	5	Q27472	Q27472 caenorhabdi
71	6	28.6	608	11	O88337	O88337 rattus norv
72	6	28.6	612	4	Q9NQX0	Q9NQX0 homo sapien
73	6	28.6	612	3	Q9US12	Q9US12 schizosacch
74	6	28.6	626	3	Q22779	Q22779 caenorhabdi
75	6	28.6	635	11	O88810	O88810 rattus norv
76	6	28.6	649	4	Q969Y5	Q969Y5 homo sapien
77	6	28.6	671	17	Q980R1	Q980R1 sulfolobus
78	6	28.6	693	5	Q9TZK7	Q9TZK7 caenorhabdi
79	6	28.6	726	11	Q9WV66	Q9WV66 mus musculu
80	6	28.6	737	16	Q9AAB8	Q9AAB8 caulobacter
81	6	28.6	787	3	Q9EYQ5	Q9EYQ5 clostridium
82	6	28.6	804	4	Q9N5V0	Q9N5V0 caenorhabdi
83	6	28.6	872	3	O13768	O13768 schizosacch
84	6	28.6	872	4	Q9H6U7	Q9H6U7 homo sapien
85	6	28.6	909	4	Q96GN3	Q96GN3 homo sapien
86	6	28.6	914	6	Q9TTT4	Q9TTT4 oryctolagus
87	6	28.6	1054	11	Q91W15	Q91W15 mus musculu
88	6	28.6	1126	10	Q9C698	Q9C698 arabidopsis
89	6	28.6		17	Q976A4	Q976A4 sulfolobus

#	cnt	val	pos	ID	n	Description
90	6	28.6	1145	Q91V56	12	Q91v56 olive laten
91	6	28.6	1184	O22725	10	O22725 arabidopsis
92	6	28.6	1263	Q9QYZ2	11	Q9qyz2 mus musculu
93	6	28.6	1291	Q9UFV0	4	Q9ufv0 homo sapien
94	6	28.6	1359	Q9UPR1	4	Q9upr1 homo sapien
95	6	28.6	1559	O73697	13	O73697 fugu rubrip
96	6	28.6	1602	Q91W25	11	Q91w25 mesocricetu
97	6	28.6	1643	Q91WX9	13	Q91wx9 rattus norv
98	6	28.6	1688	O57483	4	O57483 rana catesb
99	6	28.6	1769	Q9P273	4	Q9p273 homo sapien
100	6	28.6	1912	Q95226	4	Q95226 homo sapien
101	6	28.6	1977	Q9UHB1	4	Q9uhb1 homo sapien
102	6	28.6	1981	Q92Z7	11	Q923z7 rattus norv
103	6	28.6	1985	Q9IS7	11	Q91is7 mus musculu
104	6	28.6	2100	P87112	3	P87112 schizosacch
105	6	28.6	2168	Q902A7	13	Q902a7 brachydanio
106	6	28.6	2169	Q92305	11	Q92305 cavia porce
107	6	28.6	2271	Q99QY4	16	Q99qy4 staphylococ
108	6	28.6	2322	Q92675	4	Q92675 homo sapien
109	6	28.6	2346	Q9JIC1	11	Q9jic1 mus musculu
110	6	28.6	2590	Q9W7R4	13	Q9w7r4 brachydanio
111	6	28.6	2715	Q9WTS6	11	Q9wts6 mus musculu
112	6	28.6	3194	Q92LM3	16	Q92lm3 helicobacte
113	6	28.6	3512	O62524	5	O62524 chironomus
114	6	28.6	3972	P73139	16	P73139 synechocyst
115	6	23.8	25	Q9TRV9	6	Q9trv9 oryctolagus
116	6	23.8	40	Q9IFE0	12	Q9ife0 chilo iride
117	6	23.8	42	P70476	11	P70476 rattus norv
118	6	23.8	47	Q9JS22	16	Q9js22 neisseria m
119	6	23.8	49	Q9F214	2	Q9f214 roseateles
120	5	23.8	50	Q9DDZ2	13	Q9ddz2 micropterus
121	5	23.8	52	Q9PDF0	16	Q9pdf0 xylella fas
122	5	23.8	53	O65063	10	O65063 picea maria
123	5	23.8	53	Q9PD01	16	Q9pd01 xylella fas
124	5	23.8	54	Q23882	6	Q23882 dictyosteli
125	5	23.8	54	Q28539	6	Q28539 ovis aries
126	5	23.8	55	Q9TBJ4	8	Q9tbj4 cacomantis
127	5	23.8	56	O57575	13	O57575 cynops pyrr
128	5	23.8	57	Q96PS3	4	Q96ps3 homo sapien
129	5	23.8	57	Q39398	10	Q39398 betula verr
130	5	23.8	60	Q41933	10	Q41933 arabidopsis
131	5	23.8	61	Q23877	5	Q23877 dictyosteli
132	5	23.8	65	Q23878	5	Q23878 dictyosteli
133	5	23.8	66	Q23609	5	Q23609 caenorhabdi
134	5	23.8	67	O20149	8	O20149 chlorella v
135	5	23.8	67	Q9ER13	11	Q9er13 rattus norv
136	5	23.8	67	Q9JY08	16	Q9jy08 neisseria m
137	5	23.8	68	O48131	8	O48131 phrynosoma
138	5	23.8	70	Q98H86	16	Q98h86 rhizobium 1
139	5	23.8	74	Q9DW87	12	Q9dw87 rat cytomeg
140	5	23.8	75	Q9WUN3	11	Q9wun3 rattus norv
141	5	23.8	76	Q40946	10	Q40946 pharbitis n
142	5	23.8	76	Q40947	10	Q40947 pharbitis n
143	5	23.8	76	Q40952	10	Q40952 pharbitis n
144	5	23.8	76	Q9RRN2	16	Q9rrn2 deinococcus
145	5	23.8	77	Q9ZGZ1	5	Q9zgz1 yersinia pe
146	5	23.8	77	Q93MD7	2	Q93md7 clostridium
147	5	23.8	78	Q9YP99	11	Q9yp99 canine herp
148	5	23.8	79	Q92U6	12	Q92u6 mus musculu
149	5	23.8	79	Q9PDY5	16	Q9pdy5 xylella fas
150	5	23.8	80	Q23483	10	Q23483 arabidopsis
151	5	23.8	82	O41082	12	O41082 paramecium
152	5	23.8	83	Q23875	5	Q23875 dictyosteli
153	5	23.8	84	Q97UP8	17	Q97up8 sulfolobus
154	5	23.8	85	Q94364	5	Q94364 caenorhabdi
155	5	23.8	86	Q9B8X2	5	Q9b8x2 schistosoma
156	5	23.8	87	Q23879	5	Q23879 dictyosteli
157	5	23.8	88	Q23880	5	Q23880 dictyosteli
158	5	23.8	88	Q97588	6	Q97588 oryctolagus
159	5	23.8	88	Q9L199	5	Q9l199 arabidopsis
160	5	23.8	89	Q9S05	5	Q9s05 drosophila
161	5	23.8	89	Q9HWZ4	16	Q9hwz4 pseudomonas
162	5	23.8	90	Q977A4	17	Q977a4 sulfolobus
163	5	23.8	91	Q92LQ2	16	Q92lq2 rhizobium m
164	5	23.8	92	Q01695	3	Q01695 penicillium
165	5	23.8	92	Q9WBE3	15	Q9wbe3 human immun
166	5	23.8	94	Q56942	5	Q56942 yersinia pe
167	5	23.8	95	O28676	17	O28676 archaeoglob
168	5	23.8	96	Q90X51	13	Q90x51 fundulus he
169	5	23.8	97	Q97S59	16	Q97s59 streptococc
170	5	23.8	98	O48467	9	O48467 bacteriopha
171	5	23.8	98	Q9YWS7	15	Q9yws7 human immun
172	5	23.8	98	Q99YF9	16	Q99yf9 streptococc
173	5	23.8	99	Q9AFU5	2	Q9afu5 shigella fl
174	5	23.8	99	Q92T24	16	Q92t24 rhizobium m
175	5	23.8	100	Q9TYD9	5	Q9tyd9 styela clav
176	5	23.8	100	Q93XI7	16	Q93xi7 pharbitis n
177	5	23.8	100	Q9AJ8	16	Q9aj8 caulobacter
178	5	23.8	101	Q9PPB2	16	Q9ppb2 campylobact
179	5	23.8	103	Q05457	5	Q05457 saccharomyc
180	5	23.8	103	O15941	5	O15941 bombyx mori
181	5	23.8	103	Q9ZRW1	10	Q9zrw1 cicer ariet
182	5	23.8	103	Q71731	16	Q71731 human immun
183	5	23.8	103	Q9JSA8	16	Q9jsa8 chlamydia p
184	5	23.8	104	Q9MFC3	10	Q9mfc3 beta vulgar
185	5	23.8	104	Q39084	10	Q39084 arabidopsis
186	5	23.8	105	Q9JWB2	16	Q9jwb2 neisseria m
187	5	23.8	106	Q92GI8	16	Q92gi8 rickettsia
188	5	23.8	107	O73985	17	O73985 pyrococcus
189	5	23.8	108	O02562	5	O02562 acrolepiops
190	5	23.8	108	Q9XST8	6	Q9xst8 canis famil
191	5	23.8	109	Q9X7N0	2	Q9x7n0 streptomyce
192	5	23.8	109	P91881	5	P91881 diadromus p
193	5	23.8	109	P91877	5	P91877 diprion pin
194	5	23.8	111	Q61977	11	Q61977 mus musculu
195	5	23.8	111	Q9DY7	11	Q9dy7 mus musculu
196	5	23.8	113	Q89271	12	Q89271 carnation m
197	5	23.8	113	Q91153	12	Q91153 notophthalm
198	5	23.8	114	Q41211	10	Q41211 striga asia
199	5	23.8	114	O58083	17	O58083 pyrococcus
200	5	23.8	118	Q9YMG9	16	Q9ymg9 lymantria d
201	5	23.8	118	Q92KJ5	16	Q92kj5 rhizobium m
202	5	23.8	119	Q9XD86	2	Q9xd86 corynebacte
203	5	23.8	119	O67530	17	O67530 aquifex aeo
204	5	23.8	120	O58311	17	O58311 pyrococcus
205	5	23.8	121	P91789	5	P91789 taenia ovis
206	5	23.8	121	Q95YV8	5	Q95yv8 leishmania
207	5	23.8	122	O41658	12	O41658 influenza a
208	5	23.8	122	Q9JTV4	16	Q9jtv4 neisseria m
209	5	23.8	123	Q9VBQ4	5	Q9vbq4 drosophila
210	5	23.8	123	Q914N7	16	Q914n7 pseudomonas
211	5	23.8	124	Q9XY31	5	Q9xy31 drosophila
212	5	23.8	124	Q9XY30	5	Q9xy30 drosophila
213	5	23.8	124	Q9XY29	5	Q9xy29 drosophila
214	5	23.8	124	Q9XY28	5	Q9xy28 drosophila
215	5	23.8	124	Q9XY27	5	Q9xy27 drosophila
216	5	23.8	124	Q9XY26	5	Q9xy26 drosophila
217	5	23.8	124	Q9XY25	5	Q9xy25 drosophila
218	5	23.8	124	Q9XY24	5	Q9xy24 drosophila
219	5	23.8	124	Q9XY23	5	Q9xy23 drosophila
220	5	23.8	124	Q9XY22	5	Q9xy22 drosophila
221	5	23.8	124	Q9NIV4	5	Q9niv4 norrisia no
222	5	23.8	124	Q9DFK3	13	Q9dfk3 gillichthys
223	5	23.8	125	Q9IZY5	11	Q9izy5 mus musculu
224	5	23.8	125	Q9DDL4	13	Q9ddl4 callisaurus
225	5	23.8	125	Q9HWW9	16	Q9hww9 pseudomonas
226	5	23.8	126	Q9H7A1	4	Q9h7a1 homo sapien
227	5	23.8	128	Q929X2	16	Q929x2 listeria in
228	5	23.8	129	Q9NAY5	5	Q9nay5 papilio can
229	5	23.8	129	Q9FN32	10	Q9fn32 arabidopsis
230	5	23.8	129	Q980H5	17	Q980h5 sulfolobus
231	5	23.8	130	O57737	17	O57737 pyrococcus
232	5	23.8	131	Q07790	3	Q07790 saccharomyc
233	5	23.8	131	Q96NN2	4	Q96nn2 homo sapien
234	5	23.8	133	Q9XSB1	6	Q9xsb1 tupaia bela
235	5	23.8	133	O96104	17	O96104 apis mellif

#					Accession	Code	Description
236	5	23.8	133	16	O06717	O06717	bacillus su
237	5	23.8	134	5	068689	068689	yersinia pe
238	5	23.8	134	2	Q9VBM9	Q9VBM9	drosophila
239	5	23.8	134	11	Q99NC6	Q99NC6	rattus norv
240	5	23.8	134	11	Q99NC5	Q99NC5	mus musculu
241	5	23.8	135	4	O15203	O15203	homo sapien
242	5	23.8	136	2	Q9XC36	Q9XC36	pseudomonas
243	5	23.8	139	2	Q9F7W5	Q9F7W5	mycobacteri
244	5	23.8	139	17	O59457	O59457	pyrococcus
245	5	23.8	140	5	Q964H9	Q964H9	hydractinia
246	5	23.8	141	2	Q9L1P0	Q9L1P0	streptomyce
247	5	23.8	141	5	Q9NAY6	Q9NAY6	papilio gla
248	5	23.8	143	5	Q9BLK2	Q9BLK2	drosophila
249	5	23.8	143	5	Q9BLK0	Q9BLK0	drosophila
250	5	23.8	143	5	Q9BLK4	Q9BLK4	drosophila
251	5	23.8	143	5	Q9BLK3	Q9BLK3	drosophila
252	5	23.8	143	5	Q9BLK1	Q9BLK1	drosophila
253	5	23.8	143	5	Q9BLJ9	Q9BLJ9	drosophila
254	5	23.8	143	5	Q9BLJ8	Q9BLJ8	drosophila
255	5	23.8	143	5	Q9BLJ7	Q9BLJ7	drosophila
256	5	23.8	143	10	Q9FS32	Q9FS32	helianthus
257	5	23.8	144	2	Q9RI18	Q9RI18	yersinia pe
258	5	23.8	145	11	Q9D1I6	Q9D1I6	mus musculu
259	5	23.8	146	2	Q9LBR4	Q9LBR4	clostridium
260	5	23.8	146	10	Q41050	Q41050	pisum sativ
261	5	23.8	146	11	O88674	O88674	mus musculu
262	5	23.8	147	8	Q9MJ41	Q9MJ41	rana nigrom
263	5	23.8	147	8	Q9MJ40	Q9MJ40	rana porosa
264	5	23.8	147	8	Q9MJ39	Q9MJ39	rana porosa
265	5	23.8	147	8	Q9MJ38	Q9MJ38	rana porosa
266	5	23.8	147	8	Q9MJ37	Q9MJ37	rana porosa
267	5	23.8	147	16	Q9PIF9	Q9PIF9	campylobact
268	5	23.8	147	16	Q9KDJ9	Q9KDJ9	bacillus ha
269	5	23.8	147	16	Q9CJP9	Q9CJP9	pasteurella
270	5	23.8	148	16	Q9A2M3	Q9A2M3	caulobacter
271	5	23.8	150	2	Q52450	Q52450	pseudanabae
272	5	23.8	150	5	O18549	O18549	lytechinus
273	5	23.8	150	5	O18551	O18551	lytechinus
274	5	23.8	151	4	Q96LU1	Q96LU1	homo sapien
275	5	23.8	151	16	Q9A5N4	Q9A5N4	caulobacter
276	5	23.8	152	2	Q9F3L9	Q9F3L9	streptomyce
277	5	23.8	152	11	Q9D2A9	Q9D2A9	mus musculu
278	5	23.8	154	12	Q9I2B9	Q9I2B9	indian pean
279	5	23.8	154	16	Q9PGA7	Q9PGA7	xylella fas
280	5	23.8	154	16	Q9I4T0	Q9I4T0	pseudomonas
281	5	23.8	154	17	Q9UZB0	Q9UZB0	pyrococcus
282	5	23.8	155	4	Q9UDN5	Q9UDN5	homo sapien
283	5	23.8	155	16	Q98DB5	Q98DB5	rhizobium l
284	5	23.8	155	17	O59057	O59057	pyrococcus
285	5	23.8	156	5	Q21151	Q21151	caenorhabdi
286	5	23.8	156	11	Q99MF7	Q99MF7	rattus norv
287	5	23.8	157	2	O34417	O34417	helicobacte
288	5	23.8	157	2	Q9R300	Q9R300	helicobacte
289	5	23.8	157	2	O87748	O87748	helicobacte
290	5	23.8	157	2	O32645	O32645	helicobacte
291	5	23.8	157	2	O32646	O32646	helicobacte
292	5	23.8	157	2	O32647	O32647	helicobacte
293	5	23.8	157	2	O32648	O32648	helicobacte
294	5	23.8	157	2	O32649	O32649	helicobacte
295	5	23.8	157	2	O32650	O32650	helicobacte
296	5	23.8	157	2	O32651	O32651	helicobacte
297	5	23.8	157	2	O32652	O32652	helicobacte
298	5	23.8	157	2	O32653	O32653	helicobacte
299	5	23.8	157	2	O32654	O32654	helicobacte
300	5	23.8	157	2	O32655	O32655	helicobacte
301	5	23.8	157	16	P73080	P73080	synechocyst
302	5	23.8	157	16	Q92ID3	Q92ID3	rickettsia
303	5	23.8	157	17	O58591	O58591	pyrococcus
304	5	23.8	158	13	Q90WF5	Q90WF5	gallus gall
305	5	23.8	159	4	Q96DEI	Q96DEI	homo sapien
306	5	23.8	159	10	Q94KV2	Q94KV2	manihot esc
307	5	23.8	159	13	Q9DG95	Q9DG95	pagrus majo
308	5	23.8	159	16	Q9K6C2	Q9K6C2	bacillus ha

#					Code	Code	Description
309	5	23.8	159	16	Q98HM7	Q98hm7	rhizobium 1
310	5	23.8	159	16	Q92ZF3	Q92zf3	rhizobium m
311	5	23.8	160	5	Q9VLG0	Q9vlg0	drosophila
312	5	23.8	161	15	Q9IWT8	Q9iwt8	human immun
313	5	23.8	163	4	Q9NXW7	Q9nxw7	homo sapien
314	5	23.8	165	10	Q9SAW7	Q9saw7	triticum tu
315	5	23.8	165	10	Q9SAW6	Q9saw6	triticum ae
316	5	23.8	165	17	Q96Z25	Q96z25	sulfolobus
317	5	23.8	166	5	Q9VV19	Q9vv19	drosophila
318	5	23.8	166	16	Q97103	Q97103	clostridium
319	5	23.8	167	2	Q9RPP8	Q9rpp8	chlamydia t
320	5	23.8	167	4	Q96PG7	Q96pg7	homo sapien
321	5	23.8	167	10	Q9XFG2	Q9xfg2	arabidopsis
322	5	23.8	167	10	Q9SYX7	Q9syx7	arabidopsis
323	5	23.8	167	16	O84120	O84120	chlamydia t
324	5	23.8	168	4	Q96FU6	Q96fu6	homo sapien
325	5	23.8	168	8	O20602	O20602	uncultured
326	5	23.8	168	10	Q9LTD9	Q9ltd9	arabidopsis
327	5	23.8	169	5	Q9NFN1	Q9nfn1	schistosoma
328	5	23.8	169	15	Q9QRP1	Q9qrp1	human immun
329	5	23.8	170	5	O05136	O05136	pseudomonas
330	5	23.8	170	5	Q9NKW7	Q9nkw7	patinopecte
331	5	23.8	170	10	Q9LJ86	Q9lj86	arabidopsis
332	5	23.8	171	5	O62520	O62520	caenorhabdi
333	5	23.8	172	4	Q9NQ47	Q9nq47	homo sapien
334	5	23.8	172	4	Q96LY1	Q96ly1	homo sapien
335	5	23.8	173	4	Q96LZ3	Q96lz3	homo sapien
336	5	23.8	173	8	Q9B821	Q9b821	epipedobate
337	5	23.8	173	16	Q92XY4	Q92xy4	rhizobium m
338	5	23.8	174	5	Q9N474	Q9n474	caenorhabdi
339	5	23.8	174	10	Q942M2	Q942m2	oryza sativ
340	5	23.8	174	16	Q9PFR1	Q9pfr1	xylella fas
341	5	23.8	175	11	Q9D7R4	Q9d7r4	mus musculu
342	5	23.8	176	2	Q9WXC4	Q9wxc4	pseudomonas
343	5	23.8	176	11	Q9DC76	Q9dc76	mus musculu
344	5	23.8	177	2	Q9ADU4	Q9adu4	salmonella
345	5	23.8	177	16	Q9KSI2	Q9ksi2	vibrio chol
346	5	23.8	178	4	Q9H3V1	Q9h3v1	homo sapien
347	5	23.8	178	12	P87602	P87602	cowpox viru
348	5	23.8	180	3	Q9UTV0	Q9utv0	schizosacch
349	5	23.8	180	10	Q9S7S7	Q9s7s7	arabidopsis
350	5	23.8	180	16	Q9ZKK0	Q9zkk0	helicobacte
351	5	23.8	180	17	Q9HKQ7	Q9hkq7	thermoplasm
352	5	23.8	182	6	O77772	O77772	oryctolagus
353	5	23.8	182	15	Q85641	Q85641	moloney mur
354	5	23.8	184	10	Q9MB00	Q9mb00	brassica ca
355	5	23.8	184	16	Q9HYV2	Q9hyv2	pseudomonas
356	5	23.8	184	16	Q9A2V6	Q9a2v6	caulobacter
357	5	23.8	186	5	Q9VD65	Q9vd65	drosophila
358	5	23.8	186	10	Q9C9L1	Q9c9l1	arabidopsis
359	5	23.8	186	15	Q99F43	Q99f43	human immun
360	5	23.8	187	15	Q85640	Q85640	murine leuk
361	5	23.8	187	15	Q83401	Q83401	moloney mur
362	5	23.8	187	17	Q9YER3	Q9yer3	aeropyrum p
363	5	23.8	188	8	Q9G8U3	Q9g8u3	rhodomonas
364	5	23.8	189	1	Q9CZV1	Q9czv1	mus musculu
365	5	23.8	190	1	O31079	O31079	methanosarc
366	5	23.8	190	4	Q9UH44	Q9uh44	homo sapien
367	5	23.8	190	5	Q9XXA3	Q9xxa3	caenorhabdi
368	5	23.8	190	16	Q9RUK5	Q9ruk5	deinococcus
369	5	23.8	192	15	Q9DV18	Q9dv18	human immun
370	5	23.8	193	8	Q95SI2	Q95si2	microhyla s
371	5	23.8	193	17	O27380	O27380	methanother
372	5	23.8	195	5	Q975X4	Q975x4	sulfolobus
373	5	23.8	195	5	Q20188	Q20188	caenorhabdi
374	5	23.8	195	17	Q97XB7	Q97xb7	sulfolobus
375	5	23.8	196	2	Q9X620	Q9x620	salmonella
376	5	23.8	196	17.	O57856	O57856	pyrococcus
377	5	23.8	197	8	Q9P1S3	Q9p1s3	homo sapien
378	5	23.8	197	6	Q9BGX9	Q9bgx9	macaca fasc
379	5	23.8	197	16	Q97JH5	Q97jh5	clostridium
380	5	23.8	199	5	Q9N4S8	Q9n4s8	caenorhabdi
381	5	23.8	200	16	Q92V49	Q92v49	rhizobium m

No.			Length	Score	Accession	Description
382	23.8	5	200	17	O26617	O26617 methanother
383	23.8	5	201	16	Q9FF62	Q9FF62 arabidopsis
384	23.8	5	201	16	Q92R37	Q92r37 rhizobium m
385	23.8	5	202	10	Q94C15	Q94C15 arabidopsis
386	23.8	5	202	16	Q61533	Q61533 mus musculu
387	23.8	5	202	16	Q9JUI6	Q9JUI6 neisseria m
388	23.8	5	202	16	Q9CMK0	Q9CMK0 pasteurella
389	23.8	5	203	2	Q9WXH1	Q9WXH1 alcaligenes
390	23.8	5	203	10	Q9LVA3	Q9LVA3 arabidopsis
391	23.8	5	204	2	Q92436	Q92436 pseudomonas
392	23.8	5	204	10	Q96446	Q96446 glycine max
393	23.8	5	205	4	Q9H3V3	Q9H3v3 homo sapien
394	23.8	5	206	2	Q923X8	Q923x8 pseudomonas
395	23.8	5	206	2	Q9EWV6	Q9EWV6 streptomyce
396	23.8	5	206	16	Q9JRA2	Q9JRA2 neisseria m
397	23.8	5	207	16	Q9HXX4	Q9HXX4 pseudomonas
398	23.8	5	207	16	Q92D89	Q92D89 listeria in
399	23.8	5	210	2	Q9VGA9	Q9VGA9 drosophila
400	23.8	5	210	16	Q9KNI2	Q9KNI2 vibrio chol
401	23.8	5	210	16	Q9K6F6	Q9K6f6 bacillus ha
402	23.8	5	211	5	Q18549	Q18549 caenorhabdi
403	23.8	5	211	2	O76804	O76804 drosophila
404	23.8	5	211	5	Q9VEK7	Q9VEK7 drosophila
405	23.8	5	212	5	Q9ANH4	Q9ANH4 bradyrhizob
406	23.8	5	212	16	Q9PJS8	Q9PJS8 chlamydia m
407	23.8	5	213	13	O57503	O57503 sceloporus
408	23.8	5	214	5	Q95V17	Q95V17 frankliniel
409	23.8	5	214	10	Q9SSM9	Q9SSM9 arabidopsis
410	23.8	5	214	10	Q9M4Y0	Q9M4Y0 phalaenopsi
411	23.8	5	215	10	O81322	O81322 arabidopsis
412	23.8	5	216	2	O24739	O24739 streptomyce
413	23.8	5	216	12	Q98223	Q98223 molluscum c
414	23.8	5	216	12	Q9PBY5	Q9PBY5 xylella fas
415	23.8	5	217	15	Q90QZ7	Q90qz7 human immun
416	23.8	5	217	17	O59599	O59599 pyrococcus
417	23.8	5	218	2	Q93848	Q93848 caenorhabdi
418	23.8	5	218	17	Q9Y982	Q9Y982 aeropyrum p
419	23.8	5	219	13	O93402	O93402 carassius a
420	23.8	5	220	4	Q9BY18	Q9BY18 homo sapien
421	23.8	5	222	11	Q9D8E5	Q9D8E5 mus musculu
422	23.8	5	222	10	O23951	O23951 gossypium h
423	23.8	5	222	10	Q9XH32	Q9XH32 spinacia ol
424	23.8	5	222	17	Q97WC8	Q97WC8 sulfolobus
425	23.8	5	222	17	Q96ZL7	Q96ZL7 sulfolobus
426	23.8	5	224	6	Q9MZT3	Q9MZT3 megaderma l
427	23.8	5	224	12	Q9LTS2	Q9LTS2 arabidopsis
428	23.8	5	224	12	Q96916	Q96916 human herpe
429	23.8	5	225	4	Q9HC76	Q9HC76 homo sapien
430	23.8	5	225	4	Q96PG6	Q96PG6 homo sapien
431	23.8	5	225	10	Q943A4	Q943A4 oryza sativ
432	23.8	5	225	11	Q99N05	Q99N05 mus musculu
433	23.8	5	226	5	Q9X869	Q9X869 streptomyce
434	23.8	5	226	5	Q9VLK7	Q9VLK7 drosophila
435	23.8	5	226	11	Q9ES61	Q9ES61 mus musculu
436	23.8	5	226	11	Q9D3F6	Q9D3F6 mus musculu
437	23.8	5	226	8	P74140	P74140 synechocyst
438	23.8	5	227	8	O98679	O98679 pisum sativ
439	23.8	5	228	2	O52687	O52687 shewanella
440	23.8	5	228	16	Q9CJ23	Q9CJ23 lactococcus
441	23.8	5	229	11	Q9D9F6	Q9D9F6 mus musculu
442	23.8	5	229	11	Q9EQP0	Q9EQP0 mesocricetu
443	23.8	5	229	13	O90235	O90235 ambystoma m
444	23.8	5	230	5	Q24973	Q24973 giardia lam
445	23.8	5	230	8	Q9TA24	Q9TA24 loxodonta a
446	23.8	5	230	16	Q9CMX6	Q9CMX6 pasteurella
447	23.8	5	231	3	Q9US15	Q9US15 schizosacch
448	23.8	5	232	6	O86620	O86620 streptomyce
449	23.8	5	232	6	Q95210	Q95210 oryctolagus
450	23.8	5	233	10	Q9SWQ0	Q9SWQ0 pleurochrys
451	23.8	5	233	16	Q97I83	Q97I83 clostridium
452	23.8	5	234	16	Q93T84	Q93T84 brucella ab
453	23.8	5	234	11	Q9DCX4	Q9DCX4 mus musculu
454	23.8	5	234	17	Q9V2G0	Q9V2G0 pyrococcus
455	23.8	5	234	17	Q9HSL9	Q9hsl9 halobacteri
456	23.8	5	235	6	Q28937	Q28937 sus scrofa
457	23.8	5	236	4	Q9UM71	Q9um71 homo sapien
458	23.8	5	236	4	Q96FA7	Q96fa7 homo sapien
459	23.8	5	236	10	Q9LH79	Q91h79 arabidopsis
460	23.8	5	236	16	Q9RVR0	Q9rvr0 deinococcus
461	23.8	5	236	16	Q9ABG9	Q9abg9 caulobacter
462	23.8	5	237	2	Q93RV7	Q93rv7 streptomyce
463	23.8	5	238	4	Q96AB8	Q96ab8 homo sapien
464	23.8	5	238	10	Q94IH4	Q94ih4 nannochlori
465	23.8	5	238	10	Q94IH3	Q94ih3 nannochlori
466	23.8	5	238	10	Q94IH2	Q94ih2 nannochlori
467	23.8	5	238	10	Q94IH1	Q94ih1 nannochlori
468	23.8	5	238	10	Q94IH0	Q94ih0 nannochlori
469	23.8	5	238	10	Q94IG9	Q94ig9 nannochlori
470	23.8	5	238	10	Q94IG8	Q94ig8 chlorella k
471	23.8	5	239	2	Q93DV4	Q93dv4 unidentifie
472	23.8	5	239	4	Q96JQ5	Q96jq5 homo sapien
473	23.8	5	240	5	Q9GS33	Q9gs33 drosophila
474	23.8	5	240	12	P89299	P89299 yam mosaic
475	23.8	5	241	16	Q9PBI8	Q9pbi8 xylella fas
476	23.8	5	241	16	Q9A6V6	Q9a6v6 caulobacter
477	23.8	5	242	2	Q93F79	Q93f79 xanthomonas
478	23.8	5	242	10	Q9ZU42	Q9zu42 arabidopsis
479	23.8	5	242	10	Q94OT2	Q94ot2 arabidopsis
480	23.8	5	243	5	O76178	O76178 drosophila
481	23.8	5	243	10	Q94ED9	Q94ed9 oryza sativ
482	23.8	5	244	5	Q9N2T8	Q9n2t8 caenorhabdi
483	23.8	5	244	5	Q20804	Q20804 caenorhabdi
484	23.8	5	244	5	Q963K5	Q963k5 cercomonas
485	23.8	5	244	5	Q963K4	Q963k4 cercomonas
486	23.8	5	244	5	Q963K3	Q963k3 cercomonas
487	23.8	5	244	5	Q963K0	Q963k0 cercomonas
488	23.8	5	244	10	Q94FN8	Q94fn8 chloarachn
489	23.8	5	244	10	Q94FN7	Q94fn7 lotharella
490	23.8	5	244	10	Q94FN6	Q94fn6 lotharella
491	23.8	5	244	10	Q94FN5	Q94fn5 lotharella
492	23.8	5	245	5	Q9GS44	Q9gs44 drosophila
493	23.8	5	245	5	Q9GS37	Q9gs37 drosophila
494	23.8	5	245	10	Q9FUV2	Q9fuv2 goniomonas
495	23.8	5	246	2	Q93LL8	Q93118 nostoc punc
496	23.8	5	247	5	Q9GS48	Q9gs48 drosophila
497	23.8	5	247	10	Q9LRC6	Q91rc6 nicotiana t
498	23.8	5	247	10	Q94C79	Q94c79 arabidopsis
499	23.8	5	247	15	O40167	O40167 human immun
500	23.8	5	248	4	Q936H6	Q936h6 staphylococ
501	23.8	5	248	4	Q9H2W1	Q9h2w1 homo sapien
502	23.8	5	248	5	Q9GS40	Q9gs40 drosophila
503	23.8	5	248	5	Q23373	Q23373 caenorhabdi
504	23.8	5	248	16	Q9CIF1	Q9cif1 lactococcus
505	23.8	5	249	4	Q9H2N3	Q9h2n3 homo sapien
506	23.8	5	249	5	Q9GS36	Q9gs36 drosophila
507	23.8	5	250	3	Q00017	Q00017 aspergillus
508	23.8	5	250	4	Q9BY19	Q9by19 homo sapien
509	23.8	5	250	5	Q9VUU6	Q9vuu6 drosophila
510	23.8	5	250	5	Q9NEV7	Q9nev7 caenorhabdi
511	23.8	5	250	12	Q9YMH6	Q9ymh6 lymantria d
512	23.8	5	251	16	Q9I6F4	Q9i6f4 pseudomonas
513	23.8	5	252	2	Q9ZGG8	Q9zgg8 heliobacill
514	23.8	5	252	5	Q9NVT7	Q9nvt7 homo sapien
515	23.8	5	252	5	Q9GS43	Q9gs43 drosophila
516	23.8	5	252	5	Q9GPH5	Q9gph5 callosobruc
517	23.8	5	252	5	Q95ZU0	Q95zu0 caenorhabdi
518	23.8	5	253	5	Q95069	Q95069 drosophila
519	23.8	5	253	5	Q95074	Q95074 drosophila
520	23.8	5	253	5	Q95094	Q95094 drosophila
521	23.8	5	253	5	Q95070	Q95070 drosophila
522	23.8	5	253	5	Q95072	Q95072 drosophila
523	23.8	5	253	5	Q95082	Q95082 drosophila
524	23.8	5	253	5	Q95084	Q95084 drosophila
525	23.8	5	253	5	Q95089	Q95089 drosophila
526	23.8	5	253	5	Q95090	Q95090 drosophila
527	23.8	5	253	5	Q95091	Q95091 drosophila

#			Length	N	Accession	Description
528	5	23.8	253	5	Q95096	Q95096 drosophila
529	5	23.8	253	5	Q95093	Q95093 drosophila
530	5	23.8	253	5	Q9GS39	Q9GS39 drosophila
531	5	23.8	253	16	Q9A2Z0	Q9A2Z0 caulobacter
532	5	23.8	254	5	Q9GS45	Q9GS45 drosophila
533	5	23.8	254	5	Q9GS35	Q9GS35 drosophila
534	5	23.8	254	5	Q95RM5	Q95RM5 drosophila
535	5	23.8	254	10	Q9FG62	Q9FG62 arabidopsis
536	5	23.8	255	5	Q9AL79	Q9AL79 brucella su
537	5	23.8	255	5	P91946	P91946 drosophila
538	5	23.8	255	5	Q95075	Q95075 drosophila
539	5	23.8	255	5	Q9GS46	Q9GS46 drosophila
540	5	23.8	256	5	Q9VBI3	Q9VBI3 drosophila
541	5	23.8	257	3	Q9P443	Q9P443 aspergillus
542	5	23.8	257	16	Q97N93	Q97N93 streptococc
543	5	23.8	257	16	Q92U50	Q92U50 rhizobium m
544	5	23.8	258	2	Q99PZ3	Q99PZ3 shigella fl
545	5	23.8	258	5	Q9VIT7	Q9VIT7 drosophila
546	5	23.8	258	5	Q9GS47	Q9GS47 drosophila
547	5	23.8	258	5	O45005	O45005 caenorhabdi
548	5	23.8	258	11	Q9D7K5	Q9D7K5 mus musculu
549	5	23.8	259	5	Q9GS41	Q9GS41 drosophila
550	5	23.8	259	5	Q9GS38	Q9GS38 drosophila
551	5	23.8	259	5	Q9GS34	Q9GS34 drosophila
552	5	23.8	259	5	Q9GS31	Q9GS31 drosophila
553	5	23.8	259	5	O02625	O02625 ceratitis c
554	5	23.8	259	16	Q9KDA3	Q9KDA3 bacillus ha
555	5	23.8	262	2	Q9JRK3	Q9JRK3 lactococcus
556	5	23.8	262	5	Q9VHK2	Q9VHK2 drosophila
557	5	23.8	262	16	O67451	O67451 aquifex aeo
558	5	23.8	263	5	Q9GS42	Q9GS42 drosophila
559	5	23.8	263	5	Q9GS32	Q9GS32 drosophila
560	5	23.8	263	5	Q9GS30	Q9GS30 drosophila
561	5	23.8	263	6	Q9XS60	Q9XS60 oryctolagus
562	5	23.8	263	11	Q9D0M9	Q9D0M9 mus musculu
563	5	23.8	263	12	Q89061	Q89061 variola vir
564	5	23.8	263	13	Q90254	Q90254 gallus gall
565	5	23.8	264	2	Q9X647	Q9X647 aeromonas h
566	5	23.8	264	5	Q9GUN7	Q9GUN7 caenorhabdi
567	5	23.8	264	6	Q9TT19	Q9TT19 macaca fusc
568	5	23.8	264	12	Q91S72	Q91S72 planaria as
569	5	23.8	264	12	Q91S70	Q91S70 planaria as
570	5	23.8	264	16	Q98MV9	Q98MV9 rhizobium l
571	5	23.8	265	16	P74612	P74612 synechocyst
572	5	23.8	265	17	Q9YCQ2	Q9YCQ2 aeropyrum p
573	5	23.8	265	17	Q97ZJ5	Q97ZJ5 sulfolobus
574	5	23.8	266	2	Q93T29	Q93T29 burkholderi
575	5	23.8	266	5	Q9VMF6	Q9VMF6 drosophila
576	5	23.8	266	17	O29396	O29396 archaeoglob
577	5	23.8	267	4	Q9NXJ0	Q9NXJ0 homo sapien
578	5	23.8	268	11	Q9EQZ0	Q9EQZ0 mus musculu
579	5	23.8	268	11	Q99N10	Q99N10 mus musculu
580	5	23.8	269	3	Q9C494	Q9C494 absidia bla
581	5	23.8	269	3	Q9C493	Q9C493 absidia coe
582	5	23.8	269	3	Q9C492	Q9C492 absidia cor
583	5	23.8	269	3	Q9C2Z1	Q9C2Z1 absidia gla
584	5	23.8	269	3	Q9C2Y8	Q9C2Y8 absidia rep
585	5	23.8	269	3	Q9C491	Q9C491 actinomucor
586	5	23.8	269	3	Q9C2Y7	Q9C2Y7 amylomyces
587	5	23.8	269	3	Q9C490	Q9C490 apophysomyc
588	5	23.8	269	3	Q9C2Y3	Q9C2Y3 backusella
589	5	23.8	269	3	Q9C2Y2	Q9C2Y2 backusella
590	5	23.8	269	3	Q9C2Y0	Q9C2Y0 benjaminiel
591	5	23.8	269	3	Q9C2X9	Q9C2X9 blakeslea t
592	5	23.8	269	3	Q9C2X3	Q9C2X3 chaetocladi
593	5	23.8	269	3	Q9C2W0	Q9C2W0 chaetocladi
594	5	23.8	269	3	Q9C2V9	Q9C2V9 chlamydoabs
595	5	23.8	269	3	Q9C2X2	Q9C2X2 choanephora
596	5	23.8	269	3	Q9C2V7	Q9C2V7 circinella
597	5	23.8	269	3	Q9C153	Q9C153 circinomuco
598	5	23.8	269	3	Q9C2V8	Q9C2V8 cokeromyces
599	5	23.8	269	3	Q9C2X4	Q9C2X4 cunninghame
600	5	23.8	269	3	Q9C2X1	Q9C2X1 cunninghame
601	5	23.8	269	3	Q9C2V5	Q9c2v5 dichotomocl
602	5	23.8	269	3	Q9C2V4	Q9c2v4 dicranophor
603	5	23.8	269	3	Q9C2V6	Q9c2v6 dissophora
604	5	23.8	269	3	Q9C2V2	Q9c2v2 echinospora
605	5	23.8	269	3	Q9C2V3	Q9c2v3 ellisomyces
606	5	23.8	269	3	Q9C2V1	Q9c2v1 fennellomyc
607	5	23.8	269	3	Q9C2U7	Q9c2u7 gilbertella
608	5	23.8	269	3	Q9C2V0	Q9c2v0 gongronella
609	5	23.8	269	3	Q9C2U4	Q9c2u4 halteromyce
610	5	23.8	269	3	Q9C2U5	Q9c2u5 helicostylu
611	5	23.8	269	3	Q9C2U3	Q9c2u3 hesseltinel
612	5	23.8	269	3	Q9C2U6	Q9c2u6 hyphomucor
613	5	23.8	269	3	Q9C2U2	Q9c2u2 kirkomyces
614	5	23.8	269	3	Q9C2S9	Q9c2s9 micromucor
615	5	23.8	269	3	Q9C2T7	Q9c2t7 mortierella
616	5	23.8	269	3	Q9C2T2	Q9c2t2 mortierella
617	5	23.8	269	3	Q9C2T0	Q9c2t0 mortierella
618	5	23.8	269	3	Q9C2S5	Q9c2s5 mortierella
619	5	23.8	269	3	Q9C2S4	Q9c2s4 mortierella
620	5	23.8	269	3	Q9C2T8	Q9c2t8 mucor amphi
621	5	23.8	269	3	Q9C2T6	Q9c2t6 mucor circi
622	5	23.8	269	3	Q9C2T5	Q9c2t5 mucor hiema
623	5	23.8	269	3	Q9C2T4	Q9c2t4 mucor indic
624	5	23.8	269	3	Q9C2T1	Q9c2t1 mucor muced
625	5	23.8	269	3	Q9C2S8	Q9c2s8 mucor muced
626	5	23.8	269	3	Q9C2S7	Q9c2s7 mucor ramos
627	5	23.8	269	3	Q9C2S6	Q9c2s6 mucor recur
628	5	23.8	269	3	Q9C2T9	Q9c2t9 mycotypha a
629	5	23.8	269	3	Q9C2T3	Q9c2t3 mycotypha m
630	5	23.8	269	3	Q9C200	Q9c200 parasitella
631	5	23.8	269	3	Q9C205	Q9c205 phascolomyc
632	5	23.8	269	3	Q9C204	Q9c204 phycomyces
633	5	23.8	269	3	Q9C207	Q9c207 pilaira ano
634	5	23.8	269	3	Q9C1Z0	Q9c1z0 pilobolus u
635	5	23.8	269	3	Q9C203	Q9c203 pirella cir
636	5	23.8	269	3	Q9C202	Q9c202 poitrasia c
637	5	23.8	269	3	Q9C201	Q9c201 protomycoci
638	5	23.8	269	3	Q9C1Y2	Q9c1y2 radiomyces
639	5	23.8	269	3	Q9C1Y8	Q9c1y8 rhizomucor
640	5	23.8	269	3	Q9C1Y3	Q9c1y3 rhizomucor
641	5	23.8	269	3	Q9C1Y0	Q9c1y0 rhizomucor
642	5	23.8	269	3	Q9C1Y9	Q9c1y9 rhizopus az
643	5	23.8	269	3	Q9C1Y7	Q9c1y7 rhizopus mi
644	5	23.8	269	3	Q9C1Y6	Q9c1y6 rhizopus mi
645	5	23.8	269	3	Q9C1Y5	Q9c1y5 rhizopus ol
646	5	23.8	269	3	Q9C1Y4	Q9c1y4 rhizopus or
647	5	23.8	269	3	Q9C1V8	Q9c1v8 saksenaea v
648	5	23.8	269	3	Q9C1X9	Q9c1x9 spinellus f
649	5	23.8	269	3	Q9C1V9	Q9c1v9 sporodiniel
650	5	23.8	269	3	Q9C1W0	Q9c1w0 syncephalas
651	5	23.8	269	3	Q9C1X7	Q9c1x7 syncephalas
652	5	23.8	269	3	Q9C1X8	Q9c1x8 syzygites m
653	5	23.8	269	3	Q9C1V5	Q9c1v5 thamnidium
654	5	23.8	269	3	Q9C1V2	Q9c1v2 thamnostylu
655	5	23.8	269	3	Q9C1V3	Q9c1v3 thermomucor
656	5	23.8	269	3	Q9C1V0	Q9c1v0 umbelopsis
657	5	23.8	269	3	Q9C1U9	Q9c1u9 umbelopsis
658	5	23.8	269	3	Q9C1V1	Q9c1v1 utharomyces
659	5	23.8	269	3	Q9C1U2	Q9c1u2 zychaea mex
660	5	23.8	269	3	Q9C1U3	Q9c1u3 zygorhynchu
661	5	23.8	269	4	96HE5	96He5 homo sapien
662	5	23.8	269	16	P73457	P73457 synechocyst
663	5	23.8	269	16	Q97IK6	Q97ik6 clostridium
664	5	23.8	270	16	Q9PLT5	Q9plt5 chlamydia m
665	5	23.8	270	16	Q9I5B9	Q9i5b9 pseudomonas
666	5	23.8	271	5	O45290	O45290 caenorhabdi
667	5	23.8	271	16	084648	084648 chlamydia t
668	5	23.8	271	16	Q9CCZ2	Q9ccz2 mycobacteri
669	5	23.8	272	11	Q99JS1	Q99js1 mus musculu
670	5	23.8	272	16	083660	083660 treponema p
671	5	23.8	273	2	Q9RAG8	Q9rag8 nostoc sp.
672	5	23.8	273	3	Q96WY6	Q96wy6 ophiostoma
673	5	23.8	273	3	Q96WY5	Q96wy5 ophiostoma

#					ID	Name
674	5	23.8	273	10	O23571	O23571 arabidopsis
675	5	23.8	273	10	Q9AY77	Q9AY77 oryza sativ
676	5	23.8	273	13	Q98TD6	Q98TD6 paralichthy
677	5	23.8	274	2	Q9R9K6	Q9R9K6 paracoccus
678	5	23.8	274	5	O02624	O02624 chymomyza a
679	5	23.8	274	5	O02629	O02629 chymomyza p
680	5	23.8	274	5	O02631	O02631 drosophila
681	5	23.8	274	5	O02635	O02635 zaprionus t
682	5	23.8	274	5	O02630	O02630 drosophila
683	5	23.8	274	5	Q95080	Q95080 drosophila
684	5	23.8	274	5	O76807	O76807 drosophila
685	5	23.8	274	5	O02633	O02633 drosophila
686	5	23.8	274	5	Q9NAH7	Q9NAH7 caenorhabdi
687	5	23.8	275	2	Q9L332	Q9L332 serratia ma
688	5	23.8	275	16	Q55193	Q55193 synechocyst
689	5	23.8	276	2	Q9RE15	Q9RE15 klebsiella
690	5	23.8	276	5	Q9W5C7	Q9W5C7 drosophila
691	5	23.8	276	8	Q951B9	Q951B9 tetrahymena
692	5	23.8	276	10	Q9SDY0	Q9SDY0 lycopersico
693	5	23.8	276	13	Q9W674	Q9W674 xenopus lae
694	5	23.8	278	4	Q15509	Q15509 homo sapien
695	5	23.8	278	11	Q921K1	Q921k1 mus musculu
696	5	23.8	278	13	Q9PT89	Q9pt89 gallus gall
697	5	23.8	278	17	O28227	O28227 archaeoglob
698	5	23.8	279	5	Q962P0	Q962p0 schistosoma
699	5	23.8	279	10	Q9SDX9	Q9SDX9 lycopersico
700	5	23.8	280	5	Q9XXG9	Q9xxg9 caenorhabdi
701	5	23.8	280	10	Q9LIP8	Q9lip8 arabidopsis
702	5	23.8	280	10	Q96405	Q96405 chlorella v
703	5	23.8	280	13	Q91152	Q91152 notophthalm
704	5	23.8	280	16	Q92W65	Q92w65 rhizobium m
705	5	23.8	281	13	Q91118	Q91118 morone saxa
706	5	23.8	281	13	O90211	O90211 sparus aura
707	5	23.8	281	16	Q9CKS5	Q9CKS5 pasteurella
708	5	23.8	282	2	O87877	O87877 thauera aro
709	5	23.8	282	2	Q46980	Q46980 escherichia
710	5	23.8	282	10	Q9LJZ4	Q9LJZ4 arabidopsis
711	5	23.8	283	2	Q9RPU7	Q9rpu7 salmonella
712	5	23.8	283	3	Q06338	Q06338 saccharomyc
713	5	23.8	284	5	Q9VBD0	Q9vbd0 drosophila
714	5	23.8	285	5	Q9U4H5	Q9u4h5 drosophila
715	5	23.8	285	4	Q9R4T3	Q9h4t3 homo sapien
716	5	23.8	285	11	Q9ET48	Q9et48 marmota mon
717	5	23.8	285	16	Q98HR2	Q98hr2 rhizobium l
718	5	23.8	286	5	Q9RKD4	Q9rkd4 streptomyce
719	5	23.8	286	5	Q9VER6	Q9ver6 drosophila
720	5	23.8	286	16	O45565	O45565 caenorhabdi
721	5	23.8	286	16	Q986C4	Q986c4 rhizobium l
722	5	23.8	287	2	P94923	P94923 mycobacteri
723	5	23.8	287	2	Q93A27	Q93a27 pseudomonas
724	5	23.8	287	10	Q92V50	Q92v50 arabidopsis
725	5	23.8	287	16	O53755	O53755 mycobacteri
726	5	23.8	287	16	Q9CJW6	Q9cjw6 pasteurella
727	5	23.8	288	2	O07487	O07487 yersinia en
728	5	23.8	288	8	Q9BBY5	Q9bby5 fasciola he
729	5	23.8	288	16	Q99ZG1	Q99zg1 streptococc
730	5	23.8	288	17	O58132	O58132 pyrococcus
731	5	23.8	291	5	Q9VG03	Q9vg03 drosophila
732	5	23.8	292	5	O16260	O16260 caenorhabdi
733	5	23.8	292	2	Q9ZBW3	Q9zbw3 streptomyce
734	5	23.8	292	2	Q9GNF2	Q9gnf2 dictyosteli
735	5	23.8	292	5	Q9BML9	Q9bml9 helicospori
736	5	23.8	292	16	Q99YQ9	Q99yq9 streptococc
737	5	23.8	294	2	O31094	O31094 rhizobium l
738	5	23.8	294	2	Q9RMF3	Q9rmf3 zymomonas m
739	5	23.8	294	2	O05699	O05699 rhizobium l
740	5	23.8	294	11	Q9CPV8	Q9cpv8 mus musculu
741	5	23.8	294	16	Q9RY39	Q9ry39 deinococcus
742	5	23.8	294	16	Q9JUN8	Q9jun8 neisseria m
743	5	23.8	294	16	Q98K24	Q98k24 rhizobium l
744	5	23.8	294	16	Q97S23	Q97s23 streptococc
745	5	23.8	295	2	O30530	O30530 pseudomonas
746	5	23.8	295	2	O50280	O50280 pseudomonas
747	5	23.8	295	10	Q94BX8	Q94bx8 arabidopsis
748	5	23.8	295	11	O35248	O35248 cricetidae
749	5	23.8	295	16	Q9I323	Q9i323 pseudomonas
750	5	23.8	295	16	Q9CB40	Q9cb40 mycobacteri
751	5	23.8	295	17	Q974J0	Q97410 sulfolobus
752	5	23.8	296	16	Q9CUR8	Q9cur8 mus musculu
753	5	23.8	296	16	Q92SG7	Q92sg7 rhizobium m
754	5	23.8	297	12	Q90K06	Q9qk06 human enter
755	5	23.8	297	16	Q9HYH2	Q9hyh2 pseudomonas
756	5	23.8	297	16	Q92L10	Q92l10 rhizobium m
757	5	23.8	298	16	Q9KD33	Q9kd33 bacillus ha
758	5	23.8	299	10	Q94AY8	Q94ay8 arabidopsis
759	5	23.8	300	2	Q9RJ57	Q9rj57 streptomyce
760	5	23.8	300	2	Q9L2H6	Q9l2h6 streptomyce
761	5	23.8	300	4	Q96HZ3	Q96hz3 homo sapien
762	5	23.8	300	12	Q91LC5	Q91lc5 white spot
763	5	23.8	300	16	Q919R6	Q919r6 white spot
764	5	23.8	300	16	Q92JW4	Q92jw4 rhizobium m
765	5	23.8	301	16	O33234	O33234 mycobacteri
766	5	23.8	301	16	Q98I74	Q98i74 rhizobium l
767	5	23.8	302	16	Q92QV0	Q92qv0 rhizobium m
768	5	23.8	302	16	P72399	P72399 streptomyce
769	5	23.8	302	16	Q98DH2	Q98dh2 rhizobium l
770	5	23.8	303	2	O30476	O30476 bacillus su
771	5	23.8	303	2	Q9RCS9	Q9rcs9 synechococc
772	5	23.8	303	2	Q9KZ08	Q9kz08 streptomyce
773	5	23.8	303	5	Q93EA6	Q93ea6 rhizobium l
774	5	23.8	303	5	Q9N6G9	Q9n6g9 leishmania
775	5	23.8	303	16	O34543	O34543 bacillus su
776	5	23.8	303	16	Q97G32	Q97g32 clostridium
777	5	23.8	304	5	O24110	O24110 drosophila
778	5	23.8	305	2	O87967	O87967 streptomyce
779	5	23.8	305	11	Q9CWX0	Q9cwx0 mus musculu
780	5	23.8	305	16	Q9K9V6	Q9k9v6 bacillus ha
781	5	23.8	305	16	Q9JXM9	Q9jxm9 neisseria m
782	5	23.8	305	16	Q9JW94	Q9jw94 neisseria m
783	5	23.8	305	16	Q97Q32	Q97q32 streptococc
784	5	23.8	306	2	Q54095	Q54095 saccharopol
785	5	23.8	306	2	Q9XCL7	Q9xcl7 pseudomonas
786	5	23.8	306	5	Q9AGZ6	Q9agz6 pseudomonas
787	5	23.8	306	5	O18513	O18513 heliocidarl
788	5	23.8	306	5	Q93253	Q93253 caenorhabdi
789	5	23.8	306	10	Q9FTW9	Q9ftw9 oryza sativ
790	5	23.8	306	16	Q9HZ90	Q9hz90 pseudomonas
791	5	23.8	306	16	Q9HY52	Q9hy52 pseudomonas
792	5	23.8	306	16	Q98CI1	Q98ci1 rhizobium l
793	5	23.8	306	15	Q984Y9	Q984y9 rhizobium l
794	5	23.8	307	5	Q9GPK2	Q9gpk2 heterodera
795	5	23.8	307	16	Q23531	Q23531 caenorhabdi
796	5	23.8	307	16	Q9KKP7	Q9kkp7 vibrio chol
797	5	23.8	307	16	Q9HWQ7	Q9hwq7 pseudomonas
798	5	23.8	309	16	Q9RSO4	Q9rso4 deinococcus
799	5	23.8	311	16	Q9VDY7	Q9vdy7 drosophila
800	5	23.8	311	10	Q9LNY7	Q9lny7 arabidopsis
801	5	23.8	311	16	Q99ZG2	Q99zg2 streptococc
802	5	23.8	312	16	O62365	O62365 caenorhabdi
803	5	23.8	312	16	Q9K9D3	Q9k9d3 bacillus ha
804	5	23.8	314	5	Q18860	Q18860 caenorhabdi
805	5	23.8	314	5	Q9U3R4	Q9u3r4 caenorhabdi
806	5	23.8	314	5	Q9TY50	Q9ty50 caenorhabdi
807	5	23.8	315	16	Q9KE32	Q9ke32 bacillus ha
808	5	23.8	315	16	Q92VB3	Q92vb3 rhizobium m
809	5	23.8	316	4	Q9Y4R6	Q9y4r6 homo sapien
810	5	23.8	316	16	O67559	O67559 aquifex aeo
811	5	23.8	316	16	Q9PQI4	Q9pqi4 ureaplasma
812	5	23.8	316	16	Q9JVR1	Q9jvr1 neisseria m
813	5	23.8	316	4	Q9BI52	Q9bi52 rhizobium l
814	5	23.8	317	4	Q9NZR7	Q9nzr7 homo sapien
815	5	23.8	317	5	Q9VBH7	Q9vbh7 drosophila
816	5	23.8	317	6	O46546	O46546 mustela put
817	5	23.8	317	16	P72026	P72026 mycobacteri
818	5	23.8	317	16	Q92YJ0	Q92yj0 rhizobium m
819	5	23.8	319	5	O45006	O45006 caenorhabdi

us-09-833-017b-4.oligo.rspt

Wed Nov 13 14:19:10 2002

No.	Score		Len	N	Accession	Entry	Organism
820	23.8	5	319	6	Q9TV79	Q9tv79	oryctolagus
821	23.8	5	320	2	Q9WVX3	Q9wvx3	pediococcus
822	23.8	5	320	6	Q9XS29	Q9xs29	oryctolagus
823	23.8	5	320	16	Q9P9W7	Q9p9w7	xylella fas
824	23.8	5	320	16	Q9CP66	Q9cp66	pasteurella
825	23.8	5	321	16	Q9KLI8	Q9kli8	vibrio chol
826	23.8	5	322	2	Q46003	Q46003	campylobact
827	23.8	5	322	5	Q9XW39	Q9xw39	caenorhabdi
828	23.8	5	322	5	Q22524	Q22524	caenorhabdi
829	23.8	5	322	10	Q9FU82	Q9fu82	oryza sativ
830	23.8	5	322	10	Q9ZV49	Q9zv49	arabidopsis
831	23.8	5	323	16	Q91005	Q91005	pseudomonas
832	23.8	5	324	2	O86595	O86595	streptomyce
833	23.8	5	324	3	Q93CP3	Q93cp3	photorhabdu
834	23.8	5	324	10	Q9LZL7	Q9lzl7	arabidopsis
835	23.8	5	325	10	Q9M7D2	Q9m7d2	pisum sativ
836	23.8	5	326	3	Q9P348	Q9p348	isatchenki
837	23.8	5	326	3	Q9P337	Q9p337	pichia guil
838	23.8	5	326	3	Q9P318	Q9p318	metschnikow
839	23.8	5	326	3	Q9P317	Q9p317	clavispora
840	23.8	5	326	3	Q9P410	Q9p410	candida alb
841	23.8	5	326	3	Q9P409	Q9p409	candida dub
842	23.8	5	326	3	Q9P3Z9	Q9p3z9	candida tro
843	23.8	5	326	3	Q9P402	Q9p402	candida mal
844	23.8	5	326	3	Q9P3Z7	Q9p3z7	candida vis
845	23.8	5	326	3	Q9P400	Q9p400	candida par
846	23.8	5	326	3	Q9P401	Q9p401	clavispora
847	23.8	5	326	3	Q9P3Z4	Q9p3z4	debaryomyce
848	23.8	5	326	3	Q9P3Z6	Q9p3z6	candida zey
849	23.8	5	326	3	Q9P3X4	Q9p3x4	pichia ferm
850	23.8	5	326	3	Q9P3X3	Q9p3x3	pichia ferm
851	23.8	5	326	3	Q9P3X2	Q9p3x2	candida nor
852	23.8	5	326	3	Q9P3X1	Q9p3x1	candida nor
853	23.8	5	326	3	Q9P3X0	Q9p3x0	pichia memb
854	23.8	5	326	3	Q9P3W9	Q9p3w9	pichia memb
855	23.8	5	326	3	Q9HGQ9	Q9hgq9	pichia jadi
856	23.8	5	326	3	Q9HGQ8	Q9hgq8	pichia jadi
857	23.8	5	326	3	Q9P3W8	Q9p3w8	saccharomyc
858	23.8	5	326	12	Q07670	Q07670	human rotav
859	23.8	5	326	12	Q07697	Q07697	human rotav
860	23.8	5	326	12	Q9Q2P5	Q9q2p5	human rotav
861	23.8	5	326	16	O84381	O84381	chlamydia t
862	23.8	5	326	11	Q9PK18	Q9pk18	chlamydia m
863	23.8	5	327	11	Q9EQW3	Q9eqw3	mus musculu
864	23.8	5	327	13	Q9YH10	Q9yh10	cyprinus ca
865	23.8	5	328	13	Q9W7M5	Q9w7m5	brachydanio
866	23.8	5	328	16	Q98HQ2	Q98hq2	rhizobium l
867	23.8	5	329	5	Q95RW6	Q95rw6	drosophila
868	23.8	5	329	16	Q9RV92	Q9rv92	deinococcus
869	23.8	5	329	16	Q92DP6	Q92dp6	listeria in
870	23.8	5	330	11	Q95Y53	Q95y53	caenorhabdi
871	23.8	5	330	11	Q9CTQ7	Q9ctq7	mus musculu
872	23.8	5	330	16	Q9CNS9	Q9cns9	pasteurella
873	23.8	5	331	4	Q9H7H3	Q9h7h3	homo sapien
874	23.8	5	331	13	Q91037	Q91037	gadus morhu
875	23.8	5	332	5	Q9N466	Q9n466	caenorhabdi
876	23.8	5	332	5	O45397	O45397	caenorhabdi
877	23.8	5	332	5	Q95ZL1	Q95zl1	caenorhabdi
878	23.8	5	332	10	O82438	O82438	daucus caro
879	23.8	5	332	10	Q22963	Q22963	arabidopsis
880	23.8	5	332	11	Q9EQW2	Q9eqw2	mus musculu
881	23.8	5	332	16	Q98BX9	Q98bx9	rhizobium l
882	23.8	5	333	4	Q9H2L2	Q9h2l2	homo sapien
883	23.8	5	333	4	Q92972	Q92972	homo sapien
884	23.8	5	333	5	Q95RW7	Q95rw7	drosophila
885	23.8	5	333	16	Q9X0B0	Q9x0b0	thermotoga
886	23.8	5	334	5	Q9U5I2	Q9u5i2	artemia par
887	23.8	5	334	16	Q9K002	Q9k002	neisseria m
888	23.8	5	334	16	Q9JQX8	Q9jqx8	neisseria m
889	23.8	5	334	16	Q9HWJ0	Q9hwj0	pseudomonas
890	23.8	5	334	16	Q92WC4	Q92wc4	rhizobium m
891	23.8	5	335	2	Q93AE2	Q93ae2	neisseria m
892	23.8	5	335	5	Q9VC78	Q9vc78	drosophila
893	23.8	5	335	16	Q9S361	Q9s361	neisseria m
894	23.8	5	335	16	Q9JUV8	Q9juv8	neisseria m
895	23.8	5	335	16	Q97R07	Q97r07	streptococc
896	23.8	5	336	10	O23775	O23775	cycas revol
897	23.8	5	336	10	O23776	O23776	cycas revol
898	23.8	5	336	10	O24270	O24270	podocarpus
899	23.8	5	336	10	P93633	P93633	zea mays (m
900	23.8	5	336	10	P93634	P93634	zea mays (m
901	23.8	5	336	10	P93635	P93635	zea mays (m
902	23.8	5	336	10	P93636	P93636	zea mays (m
903	23.8	5	336	10	P93637	P93637	zea mays (m
904	23.8	5	336	10	P93638	P93638	zea mays (m
905	23.8	5	336	10	P93639	P93639	zea mays (m
906	23.8	5	336	10	Q96438	Q96438	glycine max
907	23.8	5	336	10	Q96439	Q96439	glycine max
908	23.8	5	336	10	Q96440	Q96440	glycine max
909	23.8	5	336	10	Q96441	Q96441	glycine max
910	23.8	5	336	10	Q96442	Q96442	glycine max
911	23.8	5	336	10	Q96444	Q96444	glycine max
912	23.8	5	336	10	Q96447	Q96447	glycine max
913	23.8	5	336	10	Q96448	Q96448	glycine max
914	23.8	5	336	10	Q96445	Q96445	glycine max
915	23.8	5	336	11	Q61264	Q61264	mus musculu
916	23.8	5	336	11	Q61272	Q61272	mus musculu
917	23.8	5	337	5	Q9NGU3	Q9ngu3	leishmania
918	23.8	5	337	11	Q9D865	Q9d865	mus musculu
919	23.8	5	337	11	Q9D7F4	Q9d7f4	mus musculu
920	23.8	5	337	11	Q9D775	Q9d775	mus musculu
921	23.8	5	337	11	Q9D745	Q9d745	mus musculu
922	23.8	5	337	11	Q9D6S4	Q9d6s4	mus musculu
923	23.8	5	337	11	Q9CQ55	Q9cq55	mus musculu
924	23.8	5	337	13	Q90WC5	Q90wc5	oncorhynchu
925	23.8	5	338	5	Q93K30	Q93k30	lactobacill
926	23.8	5	338	6	Q9N1T8	Q9n1t8	canis famil
927	23.8	5	338	10	Q9FMK6	Q9fmk6	arabidopsis
928	23.8	5	339	2	Q9JN80	Q9jn80	streptomyce
929	23.8	5	339	2	Q46047	Q46047	coxiella bu
930	23.8	5	339	5	Q9TW82	Q9tw82	caenorhabdi
931	23.8	5	339	5	Q27282	Q27282	caenorhabdi
932	23.8	5	339	5	O01855	O01855	caenorhabdi
933	23.8	5	339	5	O17521	O17521	caenorhabdi
934	23.8	5	339	5	O17225	O17225	caenorhabdi
935	23.8	5	339	5	O44640	O44640	caenorhabdi
936	23.8	5	339	5	Q9U1P1	Q9u1p1	caenorhabdi
937	23.8	5	339	5	Q9NAM0	Q9nam0	caenorhabdi
938	23.8	5	339	5	Q9U1V7	Q9u1v7	caenorhabdi
939	23.8	5	339	5	Q9NA47	Q9na47	caenorhabdi
940	23.8	5	339	5	Q18332	Q18332	caenorhabdi
941	23.8	5	339	5	Q21759	Q21759	caenorhabdi
942	23.8	5	339	5	Q21908	Q21908	caenorhabdi
943	23.8	5	339	5	Q18360	Q18360	caenorhabdi
944	23.8	5	339	5	Q20886	Q20886	caenorhabdi
945	23.8	5	339	5	Q18300	Q18300	caenorhabdi
946	23.8	5	339	5	Q17530	Q17530	caenorhabdi
947	23.8	5	339	5	O17746	O17746	caenorhabdi
948	23.8	5	339	5	Q9U359	Q9u359	caenorhabdi
949	23.8	5	339	5	O45860	O45860	caenorhabdi
950	23.8	5	339	5	O27363	O27363	caenorhabdi
951	23.8	5	339	5	Q966E5	Q966e5	caenorhabdi
952	23.8	5	340	5	Q95NJ0	Q95nj0	caenorhabdi
953	23.8	5	340	5	Q9BHL2	Q9bhl2	caenorhabdi
954	23.8	5	340	13	Q9I9G8	Q9i9g8	oncorhynchu
955	23.8	5	340	16	Q9PUB3	Q9pub3	anolis caro
956	23.8	5	340	16	Q9HWJ7	Q9hwj7	pseudomonas
957	23.8	5	341	2	Q9ZIS8	Q9zis8	escherichia
958	23.8	5	341	16	Q94K68	Q94k68	arabidopsis
959	23.8	5	341	16	Q92WG4	Q92wg4	rhizobium in
960	23.8	5	341	16	Q927T9	Q927t9	listeria in
961	23.8	5	342	2	O67999	O67999	bradyrhizob
962	23.8	5	342	2	Q9ZIS1	Q9zis1	escherichia
963	23.8	5	342	4	Q9NPE4	Q9npe4	homo sapien
964	23.8	5	342	5	Q9XTB6	Q9xtb6	caenorhabdi
965	23.8	5	343	2	Q9KWG8	Q9kwg8	streptococc

966	Q99QI5	5	343	2	23.8	Q9X3Q9 pseudomonas
967	Q99QI5	5	343	11	23.8	Q9JJ14 mus musculu
968	Q99QI5	5	344	5	23.8	Q9XW59 caenorhabdi
969	Q99QI5	5	345	10	23.8	Q9SWP8 pleurochrys
970	Q99QI5	5	345	16	23.8	Q9I5H8 pseudomonas
971	Q99QI5	5	346	5	23.8	Q9UAU9 caenorhabdi
972	Q99QI5	5	346	10	23.8	Q94FP0 chlorarachn
973	Q99QI5	5	346	10	23.8	Q94FN9 chlorarachn
974	Q99QI5	5	347	5	23.8	O61700 branchiosto
975	Q99QI5	5	347	5	23.8	Q9XW32 caenorhabdi
976	Q99QI5	5	347	10	23.8	Q9FUV3 cryptomonas
977	Q99QI5	5	347	10	23.8	Q9C6X9 arabidopsis
978	Q99QI5	5	348	5	23.8	Q9TZF4 caenorhabdi
979	Q99QI5	5	348	5	23.8	Q21304 caenorhabdi
980	Q99QI5	5	348	5	23.8	Q27886 caenorhabdi
981	Q99QI5	5	349	11	23.8	O61275 mus musculu
982	Q99QI5	5	350	2	23.8	Q9RKQ6 streptomyce
983	Q99QI5	5	350	5	23.8	Q27557 drosophila
984	Q99QI5	5	350	5	23.8	Q9TVM7 drosophila
985	Q99QI5	5	350	5	23.8	Q9UAY5 caenorhabdi
986	Q99QI5	5	350	10	23.8	Q9XFB2 lycopersico
987	Q99QI5	5	350	16	23.8	O07564 bacillus su
988	Q99QI5	5	351	3	23.8	O94542 schizosacch
989	Q99QI5	5	351	5	23.8	Q9LFD7 arabidopsis
990	Q99QI5	5	351	16	23.8	O53837 mycobacteri
991	Q99QI5	5	351	16	23.8	Q97HY0 clostridium
992	Q99QI5	5	352	10	23.8	Q9XFB3 lycopersico
993	Q99QI5	5	352	12	23.8	Q9IBV2 turkey herp
994	Q99QI5	5	352	12	23.8	Q9E6P5 turkey herp
995	Q99QI5	5	353	2	23.8	Q9L4Y8 vibrio chol
996	Q99QI5	5	353	5	23.8	Q9ZF09 mycoplasma
997	Q99QI5	5	353	5	23.8	Q9TYF9 drosophila
998	Q99QI5	5	353	5	23.8	Q9TYF7 drosophila
999	Q99QI5	5	353	5	23.8	Q9TVH7 drosophila
1000	Q99QI5	5	353	5	23.8	Q27608 drosophila

ALIGNMENTS

RESULT 1
Q99QI5
ID   Q99QI5   PRELIMINARY;   PRT;   46 AA.
AC   Q99QI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN PRECURSOR.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB14, H7, LT11, NG8, AND UA159;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277152; AAK01542.1; -.
DR   EMBL; AF277153; AAK01543.1; -.
DR   EMBL; AF277155; AAK01545.1; -.
DR   EMBL; AF277156; AAK01546.1; -.
DR   EMBL; AF277157; AAK01547.1; -.
DR   InterPro; IPR004288; ComC.
FT   Pfam; PF03047; ComC; 1.
SQ   CHAIN   26   46   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;   5211 MW;   38FA62B6F78FC3BF CRC64;

Query Match   100.0%; Score 21; DB 2; Length 46;
Best Local Similarity   100.0%;   Pred. No. 2.1e-15;

Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 SGSLSTFFRLFNRSFTQALGK 21
            ||||||||||||||||||||||
Db   26 SGSLSTFFRLFNRSFTQALGK 46

RESULT 2
Q9APK7
ID   Q9APK7   PRELIMINARY;   PRT;   46 AA.
AC   Q9APK7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BM71;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277151; AAK01541.1; -.
FT   CHAIN   26   46   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   46 AA;   5195 MW;   38E0B9DB5B8FC3BF CRC64;

Query Match   100.0%; Score 21; DB 2; Length 46;
Best Local Similarity   100.0%;   Pred. No. 2.1e-15;

Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 SGSLSTFFRLFNRSFTQALGK 21
            ||||||||||||||||||||||
Db   26 SGSLSTFFRLFNRSFTQALGK 46

RESULT 3
Q9APK6
ID   Q9APK6   PRELIMINARY;   PRT;   43 AA.
AC   Q9APK6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPETENCE STIMULATING PROTEIN.
GN   COMC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH1005;
RX   MEDLINE=21142515; PubMed=11208787;
RA   Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;
RT   "Natural Genetic Transformation of Streptococcus mutans Growing in
RT   Biofilms.";
RL   J. Bacteriol. 183:897-908(2001).
DR   EMBL; AF277154; AAK01544.1; -.
FT   CHAIN   26   43   COMPETENCE STIMULATING PROTEIN.
SQ   SEQUENCE   43 AA;   4927 MW;   E6A78FC3BF6156C7 CRC64;

Query Match   71.4%; Score 15; DB 2; Length 43;
Best Local Similarity   100.0%;   Pred. No. 4.9e-09;

Matches   15;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   4 LSTFFRLFNRSFTQA 18
            |||||||||||||||
Db   29 LSTFFRLFNRSFTQA 43

```
RESULT 4
Q90315
ID   Q90315      PRELIMINARY;      PRT;     253 AA.
AC   Q90315;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STEM CELL FACTOR, MEMBRANE-BOUND FORM.
OS   Coturnix coturnix (common quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=9091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96283808; PubMed=8679698;
RA   Petitte J.N., Kulik M.J.;
RT   "Cloning and characterization of cDNAs encoding two forms of avian
RT   stem cell factor.";
RL   Biochim. Biophys. Acta 1307:149-151(1996).
DR   EMBL; U43079; AAC59934.1; -.
DR   InterPro; IPR003452; SCF.
DR   Pfam; PF02404; SCF; 1.
SQ   SEQUENCE   253 AA;  28993 MW;  20709B3854F0207A CRC64;

Query Match             33.3%;  Score 7;  DB 13;  Pred. No. 8.1;
Best Local Similarity  100.0%;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
Matches    7;

Qy      7 FFRLFNR 13
            |||||||
Db    144 FFRLFNR 150


RESULT 5
Q90314
ID   Q90314      PRELIMINARY;      PRT;     287 AA.
AC   Q90314;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STEM CELL FACTOR.
OS   Coturnix coturnix (common quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=9091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96283808; PubMed=8679698;
RA   Petitte J.N., Kulik M.J.;
RT   "Cloning and characterization of cDNAs encoding two forms of avian
RT   stem cell factor.";
RL   Biochim. Biophys. Acta 1307:149-151(1996).
DR   EMBL; U43078; AAC59933.1; -.
DR   InterPro; IPR003452; SCF.
DR   Pfam; PF02404; SCF; 1.
SQ   SEQUENCE   287 AA;  32455 MW;  ABA81AEA422A702E CRC64;

Query Match             33.3%;  Score 7;  DB 13;  Pred. No. 9;
Best Local Similarity  100.0%;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
Matches    7;

Qy      7 FFRLFNR 13
            |||||||
Db    144 FFRLFNR 150


RESULT 6
Q9VR32
ID   Q9VR32      PRELIMINARY;      PRT;     929 AA.
AC   Q9VR32;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG15627 PROTEIN.
GN   CG15627.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003575; AAF50976.1; -.
DR   HSSP; P19491; 1GR2.
DR   FlyBase; FBgn0031634; CG15627.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR001320; Ion_glut_receptor.
DR   InterPro; IPR001220; Lectin_legB.
DR   InterPro; IPR001508; NMDA_receptor.
DR   InterPro; IPR001311; SBP_glut_receptor.
DR   Pfam; PF00060; lig_chan; 1.
DR   PRINTS; PR00177; NMDARECEPTOR.
DR   SMART; SM00079; PBPe; 1.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
SQ   SEQUENCE   929 AA;  106132 MW;  F4BF81C375A9CFED CRC64;

Query Match             33.3%;  Score 7;  DB 5;  Pred. No. 26;
Best Local Similarity  100.0%;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
Matches    7;

Qy     14 SFTQALG 20
            |||||||
```

```
Db    103 SFTQALG 109

RESULT 7
Q91G53
ID   Q91G53        PRELIMINARY;   PRT;   51 AA.
AC   Q91G53;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   046R.
OS   Chilo iridescent virus (CIV) (Insect iridescent virus type 6).
OC   Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
OX   NCBI_TaxID=10488;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Delius H., Darai G., Fluegel R.M.;
RT   "DNA analysis of insect iridescent virus 6: evidence for circular
RT   permutation and terminal redundancy.";
RL   J. Virol. 49:609-614(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86174607; PubMed=3959991;
RA   Lorbacher de Ruiz H., Gelderblom H., Hofmann W., Darai G.;
RT   "Insect iridescent virus type 6 induced toxic degenerative hepatitis
RT   in mice.";
RL   Med. Microbiol. Immunol. 175:43-53(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87321126; PubMed=2820141;
RA   Schnitzler P., Soltau J.B., Fischer M., Reisner H., Scholz J.,
RA   Delius H., Darai G.;
RT   "Molecular cloning and physical mapping of the genome of insect
RT   iridescent virus type 6: further evidence for circular permutation of
RT   the viral genome.";
RL   Virology 160:66-74(1987).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89073752; PubMed=3201750;
RA   Fischer M., Schnitzler P., Delius H., Darai G.;
RT   "Identification and characterization of the repetitive DNA element in
RT   the genome of insect iridescent virus type 6.";
RL   Virology 167:485-496(1988).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92196996; PubMed=1549908;
RA   Handermann M., Schnitzler P., Rosen-Wolff A., Raab K., Sonntag K.C.,
RA   Darai G.;
RT   "Identification and mapping of origins of DNA replication within the
RT   DNA sequences of the genome of insect iridescent virus type 6.";
RL   Virus Genes 6:19-32(1992).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93118242; PubMed=1475907;
RA   Sonntag K.C., Darai G.;
RT   "Characterization of the third origin of DNA replication of the genome
RT   of insect iridescent virus type 6.";
RL   Virus Genes 6:333-342(1992).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93260401; PubMed=8492091;
RA   Stohwasser R., Raab K., Schnitzler P., Janssen W., Darai G.;
RT   "Identification of the gene encoding the major capsid protein of
RT   insect iridescent virus type 6 by polymerase chain reaction.";
RL   J. Gen. Virol. 74:873-879(1993).
RN   [8]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94167241; PubMed=8121799;
RA   Schnitzler P., Hug M., Handermann M., Janssen W., Koonin E.V.,
RA   Delius H., Darai C.;
RT   "Identification of genes encoding zinc finger proteins, non-histone
RT   chromosomal HMG protein homologue, and a putative GTP phosphohydrolase
RT   in the genome of Chilo iridescent virus.";

RL   Nucleic Acids Res. 22:158-166(1994).
RN   [9]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94353641; PubMed=8073636;
RA   Sonntag K.C., Schnitzler P., Koonin E.V., Darai G.;
RT   "Chilo iridescent virus encodes a putative helicase belonging to a
RT   distinct family within the 'DEAD/H' superfamily: implications for the
RT   evolution of large DNA viruses.";
RL   Virus Genes 8:151-158(1994).
RN   [10]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95213160; PubMed=7698884;
RA   Sonntag K.C., Schnitzler P., Janssen W., Darai G.;
RT   "Identification of the primary structure and the coding capacity of
RT   the genome of insect iridescent virus type 6 between the genome
RT   coordinates 0.310 and 0.347 (7990 bp).";
RL   Intervirology 37:287-297(1994).
RN   [11]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94292906; PubMed=8021587;
RA   Schnitzler P., Sonntag K.C., Muller M., Janssen W., Bugert J.J.,
RA   Koonin E.V., Darai G.;
RT   "Insect iridescent virus type 6 encodes a polypeptide related to the
RT   largest subunit of eukaryotic RNA polymerase II.";
RL   J. Gen. Virol. 75:1557-1567(1994).
RN   [12]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98141693; PubMed=9482589;
RA   Bahr U., Tidona C.A., Darai G.;
RT   "The DNA sequence of Chilo iridescent virus between the genome
RT   coordinates 0.101 and 0.391: similarities in coding strategy between
RT   insect and vertebrate iridoviruses.";
RL   Virus Genes 15:235-245(1997).
RN   [13]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9912523; PubMed=9926400;
RA   Muller K., Tidona C.A., Bahr U., Darai G.;
RT   "Identification of a thymidylate synthase gene within the genome of
RT   Chilo iridescent virus.";
RL   Virus Genes 17:243-258(1998).
RN   [14]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99383793; PubMed=10456793;
RA   Muller K., Tidona C.A., Darai G.;
RT   "Identification of a gene cluster within the genome of Chilo
RT   iridescent virus encoding enzymes involved in viral DNA replication
RT   and processing.";
RL   Virus Genes 18:243-264(1999).
RN   [15]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21342589; PubMed=11448171;
RA   Jakob N.J., Muller K., Bahr U., Darai G.;
RT   "Analysis of the First Complete DNA Sequence of an Invertebrate
RT   Iridovirus: Coding Strategy of the Genome of Chilo Iridescent Virus.";
RL   Virology 286:182-196(2001).
RN   [16]
RP   SEQUENCE FROM N.A.
RA   Jakob N.J., Mueller K., Bahr U., Darai G.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF303741; AAK81979.1; -.
SQ   SEQUENCE   51 AA;   5875 MW;   FF3E6A1BC1B451A5 CRC64;

Query Match       28.6%;  Score 6;  DB 12;  Length 51;
Best Local Similarity  100.0%;  Pred. No. 22;
Matches     6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps     0;

Qy       3 SLSTFF 8
           ||||||
Db      35 SLSTFF 40

RESULT 8
041074
```

```
ID   O41074      PRELIMINARY;      PRT;   69 AA.
AC   O41074;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   A592R PROTEIN.
GN   A592R.
OS   Paramecium bursaria chlorella virus 1 (PBCV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX   NCBI_TaxID=10506;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98022962; PubMed=9356347;
RA   Li Y., Lu Z., Sun L., Ropp S., Kutish G.F., Rock D.L., Van Etten J.L.;
RT   "Analysis of 74 kb of DNA located at the right end of the 330-kb
RT   chlorella virus PBCV-1 genome.";
RL   Virology 237:360-377(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20013336; PubMed=10544099;
RA   Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA   Lisec A.D., Nickerson K.W., Van Etten J.L.;
RT   "Chlorella virus PBCV-1 encodes a functional homospermidine
RT   synthase.";
RL   Virology 263:254-262(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20478054; PubMed=11021991;
RA   Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT   "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT   PBCV-1.";
RL   Virology 276:27-36(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Graves M.V., Van Etten J.L.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RA   Graves M.V., Van Etten J.L.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RA   Gurnon J.R., Graves M.V., Van Etten J.L.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U42580; AAC96934.1; -.
SQ   SEQUENCE 69 AA; 7861 MW; C3F24DE5A37F79B4 CRC64;

Query Match   28.6%;  Score 6;  DB 12;  Length 69;
Best Local Similarity 100.0%;  Pred. No. 29;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    3 SLSTFF 8
         ||||||
Db   27 SLSTFF 32

RESULT 9
Q91PP4
ID   Q91PP4      PRELIMINARY;      PRT;   69 AA.
AC   Q91PP4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PRE-C PROTEIN (FRAGMENT).
GN   PRE-C.
OS   Hepatitis B virus.
OC   Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
OX   NCBI_TaxID=10407;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97;
RA   Ruifu Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97;
RA   Dai E.H., Song Y.J., Wang J., Liu H.J., Chen C.Y., Yang R.F.;
RT   "Mutation specific PCR method for detecting nt1896 mutation in HBV
RT   pre-C region.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB067454; BAB62297.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE 69 AA; 7702 MW; DAEO75CC7F0C81DE CRC64;

Query Match   28.6%;  Score 6;  DB 12;  Length 69;
Best Local Similarity 100.0%;  Pred. No. 29;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    6 TFFRLF 11
         ||||||
Db   50 TFFRLF 55

RESULT 10
Q9RU75
ID   Q9RU75      PRELIMINARY;      PRT;   122 AA.
AC   Q9RU75;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 13.8 KDA PROTEIN.
GN   DR1518.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001995; AAF11085.1; -.
DR   TIGR; DR1518; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 122 AA; 13779 MW; 0F10C1F88C99E05B CRC64;

Query Match   28.6%;  Score 6;  DB 16;  Length 122;
Best Local Similarity 100.0%;  Pred. No. 48;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 SGSLST 6
         ||||||
Db   66 SGSLST 71

RESULT 11
```

```
Q84484
ID   Q84484      PRELIMINARY;      PRT;   134 AA.
AC   Q84484;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   A164L PROTEIN.
GN   A164L.
OS   Paramecium bursaria chlorella virus 1 (PBCV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX   NCBI_TaxID=10506;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95407089; PubMed=7676624;
RA   Li Y., Lu Z., Burbank D.E., Kutish G.F., Rock D.L., Van Etten J.L.;
RT   "Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map
RT   positions 45 to 88.";
RL   Virology 212:134-150(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20013326; PubMed=10544099;
RA   Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA   Lisec A.D., Nickerson K.W., Van Etten J.L.;
RT   "Chlorella virus PBCV-1 encodes a functional homospermidine
RT   synthase.";
RL   Virology 263:254-262(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20478054; PubMed=11021991;
RA   Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT   "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT   PBCV-1.";
RL   Virology 276:27-36(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Van Etten J.L.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Graves M.V., Van Etten J.L.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RA   Graves M.V., Van Etten J.L.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RA   Gurnon J.R., Graves M.V., Van Etten J.L.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U42580; AAC96532.1; -.
SQ   SEQUENCE   134 AA;  15629 MW;  C5C5B4EF55E06551 CRC64;

Query Match             28.6%;  Score 6;  DB 12;  Length 134;
Best Local Similarity   100.0%;  Pred. No. 52;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   11 FNRSFT 16
        ||||||
Db  111 FNRSFT 116

RESULT 12
Q9BSC0
ID   Q9BSC0      PRELIMINARY;      PRT;   140 AA.
AC   Q9BSC0;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 15.2 KDA PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY, ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC005132; AAH05132.1; -.
DR   InterPro; IPR001871; bZIP.
DR   SMART; SM00338; BRLZ; 1.
KW   Hypothetical protein.
FT   NON_TER        1      1
SQ   SEQUENCE   140 AA;  15183 MW;  3A765A8AA2FF4C7D CRC64;

Query Match             28.6%;  Score 6;  DB 4;  Length 140;
Best Local Similarity   100.0%;  Pred. No. 55;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 SGSLST 6
        ||||||
Db   16 SGSLST 21

RESULT 13
Q9QRM3
ID   Q9QRM3      PRELIMINARY;      PRT;   177 AA.
AC   Q9QRM3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=96CG38.9;
RX   MEDLINE=20250533; PubMed=10791871;
RA   Bikandou B., Takehisa J., Mboudjeka I., Ido E., Kuwata T.,
RA   Miyazaki Y., Moriyama H., Harada Y., Taniguchi Y., Ichimura H.,
RA   Ikeda M., Ndolo P.J., Nzoukoudi M.Y., M'Vouenze R., M'Pandi M.,
RA   Parra H.J., M'Pele P., Hayami M.;
RT   "Genetic subtypes of HIV type 1 in Republic of Congo.";
RL   AIDS Res. Hum. Retroviruses 16:613-619(2000).
DR   EMBL; AF119215; AAD48615.1; -.
DR   InterPro; IPR000777; GP120.
DR   Pfam; PF00516; GP120; 1.
KW   AIDS; Coat protein; Glycoprotein.
FT   NON_TER        1      1
FT   NON_TER      177    177
SQ   SEQUENCE   177 AA;  20456 MW;  D7BF88A336CA5ED4 CRC64;

Query Match             28.6%;  Score 6;  DB 15;  Length 177;
Best Local Similarity   100.0%;  Pred. No. 68;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    9 RLFNRS 14
        ||||||
Db  148 RLFNRS 153

RESULT 14
Q98FX0
ID   Q98FX0      PRELIMINARY;      PRT;   180 AA.
AC   Q98FX0;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
```

```
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MLR3584 PROTEIN.
GN   MLR3584.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003002; BAB50446.1; -.
KW   Complete proteome.
SQ   SEQUENCE   180 AA;   20202 MW;  054EDA7E73943DB6 CRC64;

Query Match             28.6%;  Score 6;  DB 16;  Length 180;
Best Local Similarity   100.0%;  Pred. No. 69;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          13 RSFTQA 18
                |||||
Db          61 RSFTQA 66


RESULT 15
Q901F3
ID   Q901F3       PRELIMINARY;      PRT;    181 AA.
AC   Q901F3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=18006;
RA   Holguin A., Alvarez A., Soriano V.;
RT   "HIV-1 subtype J in Spain.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050185; AAL13161.1; -.
FT   NON_TER        1      1
FT   NON_TER      181    181
SQ   SEQUENCE   181 AA;   19870 MW;  DBD76971DEE5FC51 CRC64;

Query Match             28.6%;  Score 6;  DB 15;  Length 181;
Best Local Similarity   100.0%;  Pred. No. 69;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          10 LFNRSF 15
                |||||
Db         175 LFNRSF 180


Search completed: November  5, 2002, 11:06:58
Job time : 43.791 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2002, 20:31:42 ; Search time 830.284 Seconds
(without alignments)
736.085 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 104
Sequence: 1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum-DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105348_4834/app_query.fasta_1.398
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017_@CGN_1_1616_@runat_05112002_105348_4834 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ph:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	104	100.0	141	1	AF277151	Streptoco
2	104	100.0	141	1	AF277152	Streptoco
3	104	100.0	141	1	AF277153	Streptoco
4	104	100.0	141	1	AF277155	Streptoco
5	104	100.0	141	1	AF277156	Streptoco
6	104	100.0	141	1	AF277157	Streptoco
7	88	84.6	159	1	AF277154	Streptoco
8 c	58	55.8	186403	9	AC022558	Homo sapi
9	54	51.9	55713	9	AC073088	Homo sapi
10	54	51.9	141605	2	AC013732	Homo sapi
11 c	54	51.9	176203	2	AC094338	Rattus no
12 c	54	51.9	231995	2	AC115352	Rattus no
13	53	51.0	2613	14	WTVDI	Wound tumor
14 c	53	51.0	133226	2	AL512364	Homo sapi
15	53	51.0	135820	9	AL513497	Homo sapi
16 c	53	51.0	149216	9	AL450338	Human DNA
17	53	51.0	170203	2	AL714002	Homo sapi
18 c	53	51.0	181112	2	AL360012	Homo sapi
19 c	53	51.0	230668	2	AC121921	Mus muscu
20 c	52	50.0	10475	6	I08101	Sequence 4
21	52	50.0	10475	6	I08711	Sequence 9
22 c	52	50.0	10476	6	AX334892	Sequence
23 c	52	50.0	10476	6	AX335346	Sequence
24 c	52	50.0	10476	6	M24461	Human pulmo
25 c	52	50.0	11807	9	AF400074	Homo sapi
26 c	52	50.0	54157	8	T7A14	Arabidops
27 c	52	50.0	62001	2	AC015206	Drosophil
28 c	52	50.0	99006	2	AC096254	Rattus no
29 c	52	50.0	153094	2	AC096715	Homo sapi
30 c	52	50.0	163174	2	AC103031	Rattus no
31 c	52	50.0	164659	2	AC104143	Drosophil
32	52	50.0	169189	2	AC125293	Drosophil
33	52	50.0	185512	9	AC012454	Homo sapi
34	52	50.0	222472	2	AC026091	Homo sapi
35	52	50.0	229155	2	AC087301	Homo sapi
36	52	50.0	310909	3	AE003427	Drosophil
37 c	51.5	49.5	141659	2	AC078864	Homo sapi
38	51.5	49.5	158285	9	AC021066	Homo sapi
39	51.5	49.5	166046	2	AC080174	Homo sapi
40 c	51	49.0	104117	9	AL591591	Human DNA
41 c	51	49.0	142561	2	AC131137	Rattus no
42	51	49.0	144046	2	AC068334	Homo sapi
43 c	51	49.0	149532	2	AC123515	Oryza sat
44 c	51	49.0	151131	2	AC021945	Homo sapi
45 c	51	49.0	157581	2	AP005624	Homo sapi

ALIGNMENTS

RESULT 1

```
ORGANISM     Streptococcus mutans
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE    1 (bases 1 to 141)
AUTHORS      Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE        Natural genetic transformation of Streptococcus mutans growing in
             biofilms
JOURNAL      J. Bacteriol. 183 (3), 897-908 (2001)
MEDLINE      21142515
PUBMED       11208787
REFERENCE    2 (bases 1 to 141)
AUTHORS      Lau,P.C.Y. and Cvitkovitch,D.G.
TITLE        Direct Submission
JOURNAL      Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
             124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="GB14"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01542.1"
                     /db_xref="GI:12698430"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT       58 a      19 c      25 g      39 t
ORIGIN

Alignment Scores:
Pred. No.:              7.93e-09        Length:            141
Score:                 104.00          Matches:            21
Percent Similarity:    100.00%         Conservative:        0
Best Local Similarity: 100.00%         Mismatches:          0
Query Match:           100.00%         Indels:              0
DB:                    1               Gaps:                0

US-09-833-017B-4 (1-21) x AF277152 (1-141)

Qy    1 SerGlySerLeuSerThrPheAsnArgSerPheThrGlnAlaLeuGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCGGAAGCCTATCAACATTTTCCGGCTGTTAACAGAAGTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
          |||
Db  136 AAA 138

RESULT 3
AF277153
LOCUS       AF277153      141 bp    DNA    linear    BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain H7 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277153
VERSION     AF277153.1  GI:12698431
KEYWORDS    .
SOURCE      Streptococcus mutans.
ORGANISM    Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
AUTHORS     Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE       Natural genetic transformation of Streptococcus mutans growing in
            biofilms
JOURNAL     J. Bacteriol. 183 (3), 897-908 (2001)
```

```
AF277151
LOCUS       AF277151      141 bp    DNA    linear    BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain BM71 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277151
VERSION     AF277151.1  GI:12698427
KEYWORDS    .
SOURCE      Streptococcus mutans.
ORGANISM    Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.*
REFERENCE   1 (bases 1 to 141)
AUTHORS     Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
TITLE       Natural genetic transformation of Streptococcus mutans growing in
            biofilms
JOURNAL     J. Bacteriol. 183 (3), 897-908 (2001)
MEDLINE     21142515
PUBMED      11208787
REFERENCE   2 (bases 1 to 141)
AUTHORS     Lau,P.C.Y. and Cvitkovitch,D.G.
TITLE       Direct Submission
JOURNAL     Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="BM71"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01541.1"
                     /db_xref="GI:12698428"
                     /translation="MKKTPSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT       58 a      20 c      25 g      38 t
ORIGIN

Alignment Scores:
Pred. No.:              7.93e-09        Length:            141
Score:                 104.00          Matches:            21
Percent Similarity:    100.00%         Conservative:        0
Best Local Similarity: 100.00%         Mismatches:          0
Query Match:           100.00%         Indels:              0
DB:                    1               Gaps:                0

US-09-833-017B-4 (1-21) x AF277151 (1-141)

Qy    1 SerGlySerLeuSerThrPheArgPheAsnArgSerPheThrGlnAlaLeuGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCGGAAGCCTATCAACATTTTCCGGCTGTTAACAGAAGTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
          |||
Db  136 AAA 138

RESULT 2
AF277152
LOCUS       AF277152      141 bp    DNA    linear    BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain GB14 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277152
VERSION     AF277152.1  GI:12698429
KEYWORDS    .
SOURCE      Streptococcus mutans.
```

```
MEDLINE     21142515
PUBMED      11208787
REFERENCE   2 (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="H7"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01543.1"
                     /db_xref="GI:12698432"
                     /translation="MKKTLSLKNDFKKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a     19 c     25 g     39 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  7.93e-09        Length:         141
Percent Similarity:     104.00          Matches:         21
Best Local Similarity:  100.00%         Conservative:     0
Query Match:            100.00%         Mismatches:       0
DB:                     100.00%         Indels:           0
                        1               Gaps:             0

US-09-833-017B-4 (1-21) x AF277153 (1-141)

Qy    1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCCGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
        |||
Db  136 AAA 138

RESULT 4
AF277155
LOCUS       AF277155      141 bp      DNA      linear      BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain LT11 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277155
VERSION     AF277155.1  GI:12698435
KEYWORDS    .
SOURCE      Streptococcus mutans
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
  PUBMED    11208787
REFERENCE   2 (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
```

```
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="LT11"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01545.1"
                     /db_xref="GI:12698436"
                     /translation="MKKTLSLKNDFKKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"
BASE COUNT     58 a     19 c     25 g     39 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  7.93e-09        Length:         141
Percent Similarity:     104.00          Matches:         21
Best Local Similarity:  100.00%         Conservative:     0
Query Match:            100.00%         Mismatches:       0
DB:                     100.00%         Indels:           0
                        1               Gaps:             0

US-09-833-017B-4 (1-21) x AF277155 (1-141)

Qy    1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCCGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
        |||
Db  136 AAA 138

RESULT 5
AF277156
LOCUS       AF277156      141 bp      DNA      linear      BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain NG8 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277156
VERSION     AF277156.1  GI:12698437
KEYWORDS    .
SOURCE      Streptococcus mutans
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
  PUBMED    11208787
REFERENCE   2 (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="NG8"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
```

```
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01546.1"
                     /db_xref="GI:12698438"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT    58 a    19 c    25 g    39 t
ORIGIN

Alignment Scores:
Pred. NO.:              7.93e-09    Length:          141
Score:                 104.00      Matches:         21
Percent Similarity:    100.00%     Conservative:    0
Best Local Similarity: 100.00%     Mismatches:      0
Query Match:           100.00%     Indels:          0
DB:                    1           Gaps:            0

US-09-833-017B-4 (1-21) x AF277156 (1-141)

Qy    1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAGCTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
         |||
Db  136 AAA 138

RESULT 6
AF277157
LOCUS       AF277157    141 bp    DNA    linear    BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain UA159 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277157
VERSION     AF277157.1    GI:12698439
KEYWORDS    .
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 141)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
   PUBMED   11208787
REFERENCE   2  (bases 1 to 141)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..141
                     /organism="Streptococcus mutans"
                     /strain="UA159"
                     /db_xref="taxon:1309"
     gene            1..141
                     /gene="comC"
     CDS             1..141
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01547.1"
                     /db_xref="GI:12698440"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQAL
                     GK"
```

```
     mat_peptide     76..138
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT    58 a    19 c    25 g    39 t
ORIGIN

Alignment Scores:
Pred. NO.:              7.93e-09    Length:          141
Score:                 104.00      Matches:         21
Percent Similarity:    100.00%     Conservative:    0
Best Local Similarity: 100.00%     Mismatches:      0
Query Match:           100.00%     Indels:          0
DB:                    1           Gaps:            0

US-09-833-017B-4 (1-21) x AF277157 (1-141)

Qy    1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
         |||
Db  136 AAA 138

RESULT 7
AF277154
LOCUS       AF277154    159 bp    DNA    linear    BCT 07-FEB-2001
DEFINITION  Streptococcus mutans strain JH1005 competence stimulating protein
            precursor (comC) gene, complete cds.
ACCESSION   AF277154
VERSION     AF277154.1    GI:12698433
KEYWORDS    .
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 159)
  AUTHORS   Li,Y.H., Lau,P.C., Lee,J.H., Ellen,R.P. and Cvitkovitch,D.G.
  TITLE     Natural genetic transformation of Streptococcus mutans growing in
            biofilms
  JOURNAL   J. Bacteriol. 183 (3), 897-908 (2001)
  MEDLINE   21142515
   PUBMED   11208787
REFERENCE   2  (bases 1 to 159)
  AUTHORS   Lau,P.C.Y. and Cvitkovitch,D.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Microbiology, Dental Research Institute,
            124 Edward Street, Toronto, ON M5G 1G6, Canada
FEATURES             Location/Qualifiers
     source          1..159
                     /organism="Streptococcus mutans"
                     /strain="JH1005"
                     /db_xref="taxon:1309"
     gene            1..132
                     /gene="comC"
     CDS             1..132
                     /gene="comC"
                     /note="ComC"
                     /codon_start=1
                     /transl_table=11
                     /product="competence stimulating protein precursor"
                     /protein_id="AAK01544.1"
                     /db_xref="GI:12698434"
                     /translation="MKKTLSLKNDFKEIKTDELEIIIGGSGTLSTFFRLFNRSFTQA"
     mat_peptide     76..129
                     /gene="comC"
                     /product="competence stimulating protein"

BASE COUNT    65 a    23 c    26 g    45 t
ORIGIN

Alignment Scores:
Pred. NO.:              3.96e-06    Length:          159
Score:                 88.00       Matches:         20
```

```
Percent Similarity:      77.78%          Conservative:  1
Best Local Similarity:   74.07%          Mismatches:    0
Query Match:             84.62%          Indels:        6
DB:                           1          Gaps:          1

US-09-833-017b-4 (1-21) x AF277154 (1-159)

Qy     1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSer------------ 14
          |||||::||||||||||||||||||||||||||||||||
Db    76  AGCGAACCCTGTCAACATTTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTAGAAT 135

Qy    15  PheThrGlnAlaLeuGlyLys 21
          ||||||||||||||||||||
Db   136  TTTACACAAGCTTGGGAAAA 156
```

```
RESULT 8
AC022558/c
LOCUS       AC022558        186403 bp    DNA     linear   PRI 01-DEC-2001
DEFINITION  Homo sapiens chromosome 15, clone RP11-90B9, complete sequence.
ACCESSION   AC022558
VERSION     AC022558.9  GI:17223344
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186403)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 15, clone RP11-90B9
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 186403)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,J., Marquis,N., McEwan,P., McGurk,A., McKernan,J.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu.X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 186403)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,J., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

```
                        Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 1, 2001 this sequence version replaced gi:16974232.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
                Center project name: L2570
                Center clone name: 90_B_9
            ---------
FEATURES             Location/Qualifiers
     source          1..186403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-90B9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   2..73
                     /rpt_family="L1P"
     repeat_region   146..262
                     /rpt_family="L1M4"
     repeat_region   443..713
                     /rpt_family="L1M4"
     repeat_region   complement(806..1103)
                     /rpt_family="MER20B"
     repeat_region   complement(1196..1361)
                     /rpt_family="MER82"
     repeat_region   1383..1530
                     /rpt_family="MIR"
     repeat_region   complement(1551..1847)
                     /rpt_family="AluSq"
     repeat_region   1914..1949
                     /rpt_family="(CA)n"
     repeat_region   complement(1950..2031)
                     /rpt_family="MER82"
     repeat_region   complement(2033..2097)
                     /rpt_family="MER20B"
     repeat_region   2098..2401
                     /rpt_family="AluSp"
     repeat_region   complement(2402..2609)
                     /rpt_family="MER20B"
     repeat_region   2739..2850
                     /rpt_family="L1MB3"
     repeat_region   2851..3147
                     /rpt_family="AluY"
     repeat_region   3148..3182
                     /rpt_family="L1MB3"
     repeat_region   complement(3536..3848)
                     /rpt_family="AluSq"
     unsure          3563
                     /note="probably T"
     repeat_region   4018..4276
                     /rpt_family="AluJb"
     repeat_region   4582..4748
                     /rpt_family="L1ME"
     repeat_region   4749..5036
                     /rpt_family="AluJb"
     repeat_region   5037..5229
                     /rpt_family="L1ME"
     repeat_region   5230..5538
                     /rpt_family="AluJb"
     repeat_region   5539..5711
                     /rpt_family="L1ME"
     repeat_region   complement(5958..6237)
                     /rpt_family="AluSc"
```

```
Pred. No.:              706              Length:        186403
Score:                 58.00            Matches:       11
Percent Similarity:    93.33%           Conservative:  3
Best Local Similarity: 73.33%           Mismatches:    1
Query Match:           55.77%           Indels:        0
DB:                    9                Gaps:          0

US-09-833-017B-4 (1-21) x AC022558 (1-186403)

Qy       2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
           |||||||||||||||||||||||||||||||||||::||||||:
Db   82372 GGCTCACTCTCCTCCTTCTTTCAGTCTTTAATGCCTCCTTCTCA 82328

RESULT 9
AC073088
LOCUS        AC073088              55713 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION   Homo sapiens BAC clone RP11-323P8 from 2, complete sequence.
ACCESSION    AC073088
VERSION      AC073088.5  GI:18093125
KEYWORDS     HTG.
SOURCE       Homo sapiens.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 55713)
  AUTHORS    Sulston,J.E. and Waterston.R.
  TITLE      Toward a complete human genome sequence
  JOURNAL    Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE    99063792
   PUBMED    9847074
REFERENCE    2  (bases 1 to 55713)
  AUTHORS    Desai,A. and Abbott,A.
  TITLE      The sequence of Homo sapiens BAC clone RP11-323P8
  JOURNAL    Unpublished (2001)
REFERENCE    3  (bases 1 to 55713)
  AUTHORS    Waterston.R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-JUN-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE    4  (bases 1 to 55713)
  AUTHORS    Waterston.R.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-JAN-2002) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      On Jan 9, 2002 this sequence version replaced gi:15825633.
             --------------- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@watson.wustl.edu
             --------------- Summary Statistics
               Center project name: H_NH0323P08
             ---------------
             NOTICE: This sequence may not represent the entire insert of this
             clone. It may be shorter because we only sequence overlapping
             clone sections once, or longer because we provide a small overlap
             between neighboring data submissions.

             This sequence was finished as follows unless otherwise noted:
             all regions were double stranded, sequenced with an alternate
             chemistry, or covered by high quality data (i.e., phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by sequence
             from more than one subclone; and the assembly was confirmed by
             restriction digest.

             MAPPING INFORMATION:
             Mapping information for this clone was provided by Dr. John D.
             McPherson, Department of Genetics, Washington University, St. Louis
             MO. For additional information about the map position of this
             sequence, see http://genome.wustl.edu/gsc
```

```
repeat_region    complement(6240..6528)
                 /rpt_family="AluSx"
repeat_region    6532..6567
                 /rpt_family="AT_rich"
repeat_region    6569..6866
                 /rpt_family="AluJo"
repeat_region    6917..7015
                 /rpt_family="MIR"
repeat_region    complement(7149..7446)
                 /rpt_family="AluSx"
repeat_region    7621..7804
                 /rpt_family="L2"
repeat_region    complement(8641..8911)
                 /rpt_family="AluJo"
repeat_region    complement(8920..9052)
                 /rpt_family="AluY"
repeat_region    9053..9329
                 /rpt_family="AluSg"
unsure           9138
                 /note="probably C"
repeat_region    9431..9566
                 /rpt_family="AluJo"
repeat_region    9567..9848
                 /rpt_family="AluSx"
repeat_region    12458..12496
                 /rpt_family="AT_rich"
repeat_region    complement(12552..12849)
                 /rpt_family="AluSx"
repeat_region    complement(13140..13234)
                 /rpt_family="MIR"
repeat_region    13246..13535
                 /rpt_family="AluJo"
repeat_region    13771..13793
                 /rpt_family="AT_rich"
repeat_region    14137..14241
                 /rpt_family="FLAM_A"
repeat_region    14720..15060
                 /rpt_family="L2"
repeat_region    15214..15239
                 /rpt_family="AT_rich"
repeat_region    15284..15587
                 /rpt_family="AluSx"
repeat_region    16006..16471
                 /rpt_family="L2"
repeat_region    17776..18082
                 /rpt_family="L2"
repeat_region    19136..19177
                 /rpt_family="MIR3"
repeat_region    19594..19739
                 /rpt_family="MIR"
repeat_region    complement(20387..20680)
                 /rpt_family="AluSx"
repeat_region    complement(20712..21009)
                 /rpt_family="AluSx"
repeat_region    complement(21013..21310)
                 /rpt_family="AluSx"
repeat_region    21738..21873
                 /rpt_family="MER5B"
repeat_region    complement(22138..22441)
                 /rpt_family="AluSq"
repeat_region    22555..22575
                 /rpt_family="AT_rich"
repeat_region    complement(22666..22969)
                 /rpt_family="AluSg"
repeat_region    23336..23640
                 /rpt_family="AluSq"
repeat_region    complement(23848..24011)
                 /rpt_family="MIR3"
repeat_region    24060..24157
                 /rpt_family="MARNA"
repeat_region    24162..24192
```

Alignment Scores:

```
SOURCE INFORMATION:
  The RPCI-11 human BAC library was made from the blood of one male
  donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
  Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
  approach for construction of bacterial artificial chromosome
  libraries. Genomics 51:1-8. The clone may be obtained either from
  Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
  and coworkers at the Roswell Park Cancer Institute
  (http://bacpac.med.buffalo.edu)
  VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
  The clone sequenced to the left is RP11-207G14, 2000 bp overlap;
  the clone sequenced to the right is RP11-92K14, 2000 bp overlap.
  Actual start of this clone is at base position 67479 of
  RP11-207G14; actual end is at base position 13505 of RP11-92K14.

  Data from AC013732 was used to finish the clone, AC073088.
  Polymorphisms have been identified between AC013732 AC073088.
FEATURES             Location/Qualifiers
  source          1..55713
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="2"
                  /map="2"
                  /clone="RP11-323P8"
                  /clone_lib="RPCI-11"
  repeat_region   1..146
                  /rpt_family="L1"
  repeat_region   152..263
                  /rpt_family="MaLR"
  repeat_region   288..455
                  /rpt_family="MER1_type"
  repeat_region   506..802
                  /rpt_family="Alu"
  repeat_region   777..849
                  /rpt_family="A-rich"
  repeat_region   804..1355
                  /rpt_family="L1"
  repeat_region   2561..2602
                  /rpt_family="AT_rich"
  repeat_region   2644..3075
                  /rpt_family="L1"
  repeat_region   2683..2704
                  /rpt_family="AT_rich"
  repeat_region   2903..2948
                  /rpt_family="AT_rich"
  repeat_region   3272..4082
                  /rpt_family="L1"
  repeat_region   4083..4134
                  /rpt_family="ERV1"
  repeat_region   4135..4390
                  /rpt_family="ERV1"
  repeat_region   4392..4463
                  /rpt_family="ERV1"
  repeat_region   4464..4591
                  /rpt_family="ERV1"
  repeat_region   4592..4767
                  /rpt_family="L1"
  repeat_region   5639..5882
                  /rpt_family="MER1_type"
  repeat_region   5883..6179
                  /rpt_family="Alu"
  repeat_region   6157..6179
                  /rpt_family="(A)n"
  repeat_region   6180..6425
                  /rpt_family="MER1_type"
  repeat_region   6222..6265
                  /rpt_family="AT_rich"
  repeat_region   6281..6312
                  /rpt_family="AT_rich"
  repeat_region   6893..6944
                  /rpt_family="AT_rich"
  repeat_region   7018..7296
                  /rpt_family="Alu"
  repeat_region   7474..7760
                  /rpt_family="L1"
  repeat_region   7805..7965
                  /rpt_family="L1"
  repeat_region   7990..8247
                  /rpt_family="MaLR"
  repeat_region   8248..8516
                  /rpt_family="Alu"
  repeat_region   8517..8690
                  /rpt_family="MaLR"
  repeat_region   9638..9752
                  /rpt_family="MIR"
  repeat_region   10151..11635
                  /rpt_family="L1"
  repeat_region   12015..12071
                  /rpt_family="(TG)n"
  repeat_region   12690..13000
                  /rpt_family="Alu"
  repeat_region   12971..13000
                  /rpt_family="(A)n"
  repeat_region   13184..13266
                  /rpt_family="ERVL"
  repeat_region   14373..14504
                  /rpt_family="L1"
  repeat_region   15153..15436
                  /rpt_family="L1"
  repeat_region   15680..16605
                  /rpt_family="L1"
  repeat_region   16613..17007
                  /rpt_family="MaLR"
  repeat_region   16755..16805
                  /rpt_family="(TC)n"
  repeat_region   17008..17241
                  /rpt_family="L1"
  repeat_region   17242..17543
                  /rpt_family="Alu"
  repeat_region   17523..17543
                  /rpt_family="(A)n"
  repeat_region   17544..17670
                  /rpt_family="L1"
  repeat_region   17728..17759
                  /rpt_family="(TG)n"
  repeat_region   17800..18266
                  /rpt_family="L1"
  repeat_region   18267..18571
                  /rpt_family="Alu"
  repeat_region   18572..19590
                  /rpt_family="L1"
  repeat_region   19242..19340
                  /rpt_family="A-rich"
  repeat_region   19594..19772
                  /rpt_family="L1"
  repeat_region   19782..20196
                  /rpt_family="L1"
  repeat_region   20222..20884
                  /rpt_family="L1"
  repeat_region   20903..22061
                  /rpt_family="L1"
  misc_feature    22442..22896
                  /note="similar to EST AW991366 (NID:g8251411)"
  repeat_region   22555..22662
                  /rpt_family="T-rich"
  repeat_region   23176..25314
                  /rpt_family="L1"
  repeat_region   25328..25489
                  /rpt_family="L1"
  repeat_region   25462..25497
                  /rpt_family="Alu"

Alignment Scores:
  Pred. No.:                                    884              Length:     55713
  Score:                                        54.00            Matches:    13
```

```
Percent Similarity:       70.00%        Conservative:    1
Best Local Similarity:    65.00%        Mismatches:      6
Query Match:              51.92%        Indels:          0
DB:                       9             Gaps:            0

US-09-833-017B-4 (1-21) x AC073088 (1-55713)

Qy      2  GlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys  21
           |||||||||| |||  |||||   |||||||||||||||::::||| |||
Db  20717  GGAAGTCTATCAAACATCTAAAGAAGGTTTAACATCAGTTTACTCAGTTTACTCCTTTGAAA  20776


RESULT 10
AC013732
LOCUS        AC013732          141605 bp    DNA    linear    HTG 18-FEB-2001
DEFINITION   Homo sapiens chromosome 2 clone RP11-550O23, WORKING DRAFT
             SEQUENCE, 9 unordered pieces.
ACCESSION    AC013732
VERSION      AC013732.6  GI:9838265
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       Homo sapiens
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 141605)
  AUTHORS    Waterston,R.H.
  TITLE      The sequence of Homo sapiens clone
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 141605)
  AUTHORS    Waterston,R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-NOV-1999) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      On Aug 17, 2000 this sequence version replaced gi:8961220.

             -------------------- Genome Center --------------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc/index.shtml
             -------------- Project Information --------------
             Center project name: H_NH0550O23
             -------------- Summary Statistics --------------
             Sequencing vector: M13; 73%
             Sequencing vector: plasmid; 27%
             Chemistry: Dye-primer ET; 73% of reads
             Chemistry: Dye-terminator Big Dye; 27% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 116598 bases at least Q40
             Consensus quality: 138339 bases at least Q30
             Consensus quality: 139218 bases at least Q20
             Insert size: 144000; agarose-fp
             Insert size: 141800; sum-of-contigs
             Quality coverage: 4.32 in Q20 bases; agarose-fp
             Quality coverage: 4.85 in Q20 bases; sum-of-contigs

             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 9 contigs. The true order of the pieces
             * is not known and their order in this sequence record as
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     5272: contig of 5272 bp in length
             *    5273     5372: gap of unknown length
             *    5373     9928: contig of 4556 bp in length
             *    9929    10028: gap of unknown length
             *   10029    19924: contig of 9896 bp in length
             *   19925    20024: gap of unknown length
             *   20025    33364: contig of 13340 bp in length
             *   33365    33464: gap of unknown length
             *   33465    47072: contig of 13608 bp in length
             *   47073    47172: gap of unknown length
             *   47173    67693: contig of 20521 bp in length
             *   67694    67793: gap of 17877 bp in length
             *   67794    85670: contig of 17877 bp in length
             *   85671    85770: gap of unknown length
             *   85771   110614: contig of 24844 bp in length
             *  110615   110714: gap of unknown length
             *  110715   141605: contig of 30891 bp in length.

FEATURES             Location/Qualifiers
     source          1..141605
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-550O23"
     misc_feature    1..5272
                     /note="assembly_name:Contig6"
     misc_feature    5373..9928
                     /note="assembly_name:Contig7"
     misc_feature    10029..19924
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:right"
     misc_feature    20025..33364
                     /note="assembly_name:Contig9"
     misc_feature    33465..47072
                     /note="assembly_name:Contig10"
     misc_feature    47173..67693
                     /note="assembly_name:Contig11"
     misc_feature    67794..85670
                     /note="assembly_name:Contig12"
     misc_feature    85771..110614
                     /note="assembly_name:Contig13"
     misc_feature    110715..141605
                     /note="assembly_name:Contig14
                     clone_end:T7
                     vector_side:left"

BASE COUNT   43124 a   29167 c   28457 g   40051 t   806 others
ORIGIN

Alignment Scores:
Pred. No.:                2.41e+03      Length:         141605
Score:                    54.00         Matches:        13
Percent Similarity:       70.00%        Conservative:   1
Best Local Similarity:    65.00%        Mismatches:     6
Query Match:              51.92%        Indels:         0
DB:                       2             Gaps:           0

US-09-833-017B-4 (1-21) x AC013732 (1-141605)

Qy      2  GlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys  21
           |||||||||| |||  |||||   |||||||||||||||::::||| |||
Db  17509  GGAAGTCTATCAAACATCTAAAGAAGGTTTAACATCAGTTTACTCAATCTCTTTGAAA  17568


RESULT 11
AC094338
LOCUS        AC094338          176203 bp    DNA    linear    HTG 10-JUL-2002
DEFINITION   Rattus norvegicus clone CH230-3J13, *** SEQUENCING IN PROGRESS ***,
             62 unordered pieces.
ACCESSION    AC094338
VERSION      AC094338.3  GI:21716581
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 176203)
  AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
```

```
          Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
          Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,D.,
          Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176203)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 176203)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jul 9, 2002 this sequence version replaced gi:17941058.
COMMENT
            ---------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
               Center project name: GAKJ
               Center clone name: CH230-3J13
            ---------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 133645 bases at least Q40
               Consensus quality: 139597 bases at least Q30
               Consensus quality: 143465 bases at least Q20
            ----------
          * NOTE: Estimated insert size may differ from sequence length
          *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *       consists of 62 contigs. The true order of the pieces
          *       is not known and their order in this sequence record is
          *       arbitrary. Gaps between the contigs are represented as
          *       runs of N, but the exact sizes of the gaps are unknown.
          *       This record will be updated with the finished sequence
          *       as soon as it is available and the accession number will
          *       be preserved.
          *
          *         1   1045: contig of 1045 bp in length
          *      1046   1145: gap of unknown length
          *      1146   2720: contig of 1575 bp in length
          *      2721   2820: gap of unknown length
          *      2821   3898: contig of 1078 bp in length
          *      3899   3998: gap of unknown length
          *      3999   5208: contig of 1210 bp in length
          *      5209   5308: gap of unknown length
          *      5309   6558: contig of 1250 bp in length
          *      6559   6658: gap of unknown length
          *      6659   7907: contig of 1249 bp in length
          *      7908   8007: gap of unknown length
          *      8008   9485: contig of 1478 bp in length
          *      9486   9585: gap of unknown length
          *      9586  10608: contig of 1023 bp in length
          *     10609  10708: gap of unknown length
          *     10709  12495: contig of 1787 bp in length
          *     12496  12595: gap of unknown length
          *     12596  13654: contig of 1059 bp in length
          *     13655  13754: gap of unknown length
          *     13755  15125: contig of 1371 bp in length
          *     15126  15225: gap of unknown length
          *     15226  16288: contig of 1063 bp in length
          *     16289  16388: gap of unknown length
          *     16389  18461: contig of 2073 bp in length
          *     18462  18561: gap of unknown length
          *     18562  19771: contig of 1210 bp in length
          *     19772  19871: gap of unknown length
          *     19872  21886: contig of 2015 bp in length
          *     21887  21986: gap of unknown length
          *     21987  23857: contig of 1871 bp in length
          *     23858  23957: gap of unknown length
          *     23958  25478: contig of 1521 bp in length
          *     25479  25578: gap of unknown length
          *     25579  27590: contig of 2012 bp in length
          *     27591  27690: gap of unknown length
          *     27691  29072: contig of 1382 bp in length
          *     29073  29172: gap of unknown length
          *     29173  31414: contig of 2242 bp in length
          *     31415  31514: gap of unknown length
          *     31515  33681: contig of 2167 bp in length
          *     33682  33781: gap of unknown length
          *     33782  35801: contig of 2020 bp in length
          *     35802  35901: gap of unknown length
          *     35902  37295: contig of 1394 bp in length
          *     37296  37395: gap of unknown length
          *     37396  39262: contig of 1867 bp in length
          *     39263  39362: gap of unknown length
          *     39363  41122: contig of 1760 bp in length
          *     41123  41222: gap of unknown length
          *     41223  42968: contig of 1746 bp in length
          *     42969  43068: gap of unknown length
          *     43069  44637: contig of 1569 bp in length
          *     44638  44737: gap of unknown length
          *     44738  46698: contig of 1961 bp in length
          *     46699  46798: gap of unknown length
          *     46799  48941: contig of 2143 bp in length
          *     48942  49041: gap of unknown length
          *     49042  51116: contig of 2075 bp in length
          *     51117  51216: gap of unknown length
          *     51217  53125: contig of 1909 bp in length
          *     53126  53225: gap of unknown length
          *     53226  54938: contig of 1713 bp in length
          *     54939  55038: gap of unknown length
          *     55039  57996: contig of 2958 bp in length
          *     57997  58096: gap of unknown length
          *     58097  60022: contig of 1926 bp in length
          *     60023  60122: gap of unknown length
          *     60123  63344: contig of 3222 bp in length
          *     63345  63444: gap of unknown length
          *     63445  65874: contig of 2430 bp in length
          *     65875  65974: gap of unknown length
```

```
*    65975    67978: contig of 2004 bp in length
*    67979    68078: gap of unknown length
*    68079    71315: contig of 3237 bp in length
*    71316    71415: gap of unknown length
*    71416    73127: contig of 1712 bp in length
*    73128    73227: gap of unknown length
*    73228    76417: contig of 3190 bp in length
*    76418    76517: gap of unknown length
*    76518    78438: contig of 1921 bp in length
*    78439    78538: gap of unknown length
*    78539    80999: contig of 2461 bp in length
*    81000    81099: gap of unknown length
*    81100    83527: contig of 2428 bp in length
*    83528    83627: gap of unknown length
*    83628    86811: contig of 3184 bp in length
*    86812    86911: gap of unknown length
*    86912    90203: contig of 3292 bp in length
*    90204    90303: gap of unknown length
*    90304    93631: contig of 3328 bp in length
*    93632    93731: gap of unknown length
*    93732    97359: contig of 3628 bp in length
*    97360    97459: gap of unknown length
*    97460   102830: contig of 5371 bp in length
*   102831   102930: gap of unknown length
*   102931   106451: contig of 3521 bp in length
*   106452   106551: gap of unknown length
*   106552   110852: contig of 4301 bp in length
*   110853   110952: gap of unknown length
*   110953   113630: contig of 2678 bp in length
*   113631   113730: gap of unknown length
*   113731   117590: contig of 3860 bp in length
*   117591   117690: gap of unknown length
*   117691   122467: contig of 4777 bp in length
*   122468   122567: gap of unknown length
*   122568   127164: contig of 4597 bp in length
```

```
Alignment Scores:
Pred. No.:                3.04e+03      Length:         176203
Score:                    54.00         Matches:        11
Percent Similarity:       77.78%        Conservative:   3
Best Local Similarity:    61.11%        Mismatches:     4
Query Match:              51.92%        Indels:         0
DB:                       2             Gaps:           0
```

```
US-09-833-017B-4 (1-21) x AC094338 (1-176203)

Qy      1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAla 18
          ||||||||  |||||   |||||:||:||:|||||||:||:|||||||||::
Db 115843 TCAGGCTCTAACCCTGTTCTTTTCTCTCTTTCATCGTGCATTTACACAGTCA 115896
```

RESULT 12

```
AC115352/c
LOCUS       AC115352          231995 bp    DNA    *** linear    HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-216B10, *** SEQUENCING IN PROGRESS
            ***, 24 unordered pieces.
ACCESSION   AC115352
VERSION     AC115352.7  GI:21746758
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231995)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
```

```
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Duran-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu.X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,J., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod.B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 231995)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231995)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:20467758.
            -----------------------------------------------------------
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -----------------------------------------------------------
            Center project name: GUZB
            Center clone name: CH230-216B10
            -----------------------------------------------------------
            Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 207690 bases at least Q40
            Consensus quality: 209760 bases at least Q30
            Consensus quality: 211534 bases at least Q20
            -----------------------------------------------------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *    consists of 24 contigs. The true order of the pieces
            *    is not known and their order in this sequence record is
            *    arbitrary. Gaps between the contigs are represented as
            *    runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            *    as soon as it is available and the accession number will
            *    be preserved.
            *    1   1061: contig of 1061 bp in length
```

```
 *
 *      1062         1161: gap of unknown length
 *      1162         2485: contig of 1324 bp in length
 *      2486         2585: gap of unknown length
 *      2586         3693: contig of 1108 bp in length
 *      3694         3793: gap of unknown length
 *      3794         5268: contig of 1475 bp in length
 *      5269         5368: gap of unknown length
 *      5369         6766: contig of 1398 bp in length
 *      6767         6866: gap of unknown length
 *      6867         8147: contig of 1281 bp in length
 *      8148         8247: gap of unknown length
 *      8248         9727: contig of 1480 bp in length
 *      9728         9827: gap of unknown length
 *      9828        11062: contig of 1235 bp in length
 *     11063        11162: gap of unknown length
 *     11163        12390: contig of 1228 bp in length
 *     12391        12490: gap of unknown length
 *     12491        14165: contig of 1675 bp in length
 *     14166        14265: gap of unknown length
 *     14266        15739: contig of 1474 bp in length
 *     15740        15839: gap of unknown length
 *     15840        18806: contig of 2967 bp in length
 *     18807        18906: gap of unknown length
 *     18907        22833: contig of 3927 bp in length
 *     22834        22933: gap of unknown length
 *     22934        27301: contig of 4368 bp in length
 *     27302        27401: gap of unknown length
 *     27402        34418: contig of 7017 bp in length
 *     34419        34518: gap of unknown length
 *     34519        39960: contig of 5342 bp in length
 *     39961        39960: gap of unknown length
 *     39961        47538: contig of 7578 bp in length
 *     47539        47638: gap of unknown length
 *     47639        53621: contig of 5983 bp in length
 *     53622        53721: gap of unknown length
 *     53722        72770: contig of 19049 bp in length
 *     72771        72870: gap of unknown length
 *     72871        88310: contig of 15440 bp in length
 *     88311        88410: gap of unknown length
 *     88411       106609: contig of 18199 bp in length
 *    106610       106709: gap of unknown length
 *    106710       130355: contig of 23646 bp in length
 *    130356       130455: gap of unknown length
 *    130456       174069: contig of 43614 bp in length
 *    174070       174169: gap of unknown length
 *    174170       231995: Contig of 57826 bp in length.
```

FEATURES   Location/Qualifiers
     source   1..231995
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-216810"
BASE COUNT   67304 a 43019 c 43334 g 65782 t 12556 others
ORIGIN

US-09-833-017B-4 (1-21) x AC115352 (1-231995)

Alignment Scores:
Pred. No.:   4.09e+03
Score: 54.00   Length: 231995
Percent Similarity: 77.78%   Matches: 11
Best Local Similarity: 61.11%   Conservative: 3
Query Match: 51.92%   Mismatches: 4
DB: 2   Indels: 0
                                                 Gaps: 0

Qy   1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAla 18
           ||||||| |||||| ||||:::||:::||||||||||:::
Db 226499 TCAGGCTCCTAACCCTCGTCTTCTTTCTCTTCTTTCATCGTGCATTTACACAGTCA 226446

RESULT 13
WTVDI/c
LOCUS   WTVDI   2613 bp ds-RNA   linear   VRL 03-AUG-1993
DEFINITION Wound tumor virus genome segment S5 encoding polypeptide P5,
            complete cds.

ACCESSION   J03020
VERSION   J03020.1 GI:36169
KEYWORDS   polypeptide P5.
SOURCE   Wound tumor virus (inoculum RB) RNA, passed in Agallia constricta
            cell line AC20.
  ORGANISM Wound tumor virus
            Viruses; dsRNA viruses; Reoviridae; Phytoreovirus.
REFERENCE   1 (bases 1 to 2613)
  AUTHORS   Anzola,J.V., Xu,Z.K., Asamizu,T. and Nuss,D.L.
  TITLE   Segment-specific inverted repeats found adjacent to conserved
            terminal sequences in wound tumor virus genome and defective
            interfering RNAs
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (23), 8301-8305 (1987)
  MEDLINE   88068582
   PUBMED   3479793
COMMENT   Draft entry and printed copy of sequence for [1] kindly provided by
            D.L.Nuss, 10/13/87.
FEATURES   Location/Qualifiers
     source   1..2613
                     /organism="Wound tumor virus"
                     /db_xref="taxon:10987"
     CDS   26..2440
                     /note="polypeptide P5"
                     /codon_start=1
                     /protein_id="AAA48499.1"
                     /db_xref="GI:336170"
                     /translation="MAIDSYCIPNFSCQTIDNRTIVNIFQSCKYRSQLSVCFLNDKSAA
                     DKFSNSMRQGSGTITFIIHAEDGEISEQLHSTFRSVSTMLCGMQLFVFIVAPRNVIS
                     SETGKAITWAFRGSFIELRDHGRGEQALHDILEQFYRLSPLVNVPKMGMAYYGPTSFA
                     ELLSLSSKNKTSWRYVIDYSMFTRSALVGFASHMMDECSFANKQINVIGYNPPYVWAG
                     LRHGVTRFTEMSTPDPEGYGPIKLILPRLTGNVLLKKVKVQHDPQKKLLCDDSVMF
                     ALSRNILYIGVYPATHLLDYNLKGWRMVAVDPKINAAWAETLKQRTSIDLVPISAKFE
                     FNAQSTRDIVLKYFSGVPFSIIDDSWVEGTEDYEKFQELKQSYFEQLVMNGSTSKLRV
                     SMISMKWNRTKDVKCRRLLALLPQPYGGSLRELRAYPHVNGAAEVNIKKSEVNSYMDK
                     FTSLSISEQIGSQKFMHMLITNYGDALKLKTGRDKAIIASYSLSNAINKKERVLKFLS
                     DAAKSETLIIFGAPNLNRVKFMIKSGIVLGSDVTISNDLITFKNASGKVWKDYGYTQS
                     ELIKSSMIEITIEQMLCISSSSYNGVGYFANSIYNDMFSWYVPEWLFEKYFSIQDIRL
                     SPVALVKCFTTSIRNLCYVPHLTYYALRGSFVEKVLITNNVLNSSYLITGTSHSTFKV
                     LSNFEVPSPAGVLKFKAGDDVNISGHLLSLVIAAHFVASPLLWATHMKRMTTPVNLP
                     KNLDKLLFFDNKIKNGMLEKWHSREEVVLAAMIVENVVAHILNGRHSEIIQEITQVI
                     YEKFNA"
BASE COUNT   825 a   428 c   545 g   815 t
ORIGIN   Unreported.

Alignment Scores:
Pred. No.:   48.4
Score: 53.00   Length: 2613
Percent Similarity: 68.42%   Matches: 10
Best Local Similarity: 52.63%   Conservative: 3
Query Match: 50.96%   Mismatches: 6
DB: 14   Indels: 0
                                                 Gaps: 0

US-09-833-017B-4 (1-21) x WTVDI (1-2613)

Qy   2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
          ||||| ||| |||||||||||||:::||| :::||||||
Db 816 GGATCATGCTGTAGCTATTTACCTTCTTTAACAGGACATTTCCAGTAAGCCTAGGT 760

RESULT 14
AL512364
LOCUS   AL512364   133226 bp   DNA   linear   HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 6 clone RP11-574F19, *** SEQUENCING IN
            PROGRESS ***, 25 unordered pieces.
ACCESSION   AL512364
VERSION   AL512364.3 GI:11995201
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   Burton,J.
  AUTHORS   Direct Submission
  TITLE

```
JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk
          On Dec 28, 2000 this sequence version replaced gi:11995149.
COMMENT   ------------- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ------------- Project Information
          Center project name: bA574F19
          ------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid: L08752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 120707 bases at least Q40
          Consensus quality: 125263 bases at least Q30
          Consensus quality: 127799 bases at least Q20
          Insert size: 130826; sum-of-contigs
          Insert size: 158221; 10.3% error; agarose-fp
          Quality coverage: 3.07x in Q20 bases; sum-of-contigs Quality
          coverage: 2.92x in Q20 bases; agarose-fp
          -------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 25 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *          1   4495: contig of 4495 bp in length
          *       4496   4595: gap of    100 bp
          *       4596  17538: contig of 12943 bp in length
          *      17539  17638: gap of    100 bp
          *      17639  21198: contig of 3560 bp in length
          *      21199  21298: gap of    100 bp
          *      21299  24416: contig of 3118 bp in length
          *      24417  24516: gap of    100 bp
          *      24517  27845: contig of 3329 bp in length
          *      27846  27945: gap of    100 bp
          *      27946  35081: contig of 7136 bp in length
          *      35082  35181: gap of    100 bp
          *      35182  37225: contig of 2044 bp in length
          *      37226  37325: gap of    100 bp
          *      37326  39611: contig of 2286 bp in length
          *      39612  39711: gap of    100 bp
          *      39712  42865: contig of 3154 bp in length
          *      42866  42965: gap of    100 bp
          *      42966  49369: contig of 6404 bp in length
          *      49370  49469: gap of    100 bp
          *      49470  52200: contig of 2731 bp in length
          *      52201  52300: gap of    100 bp
          *      52301  55073: contig of 2773 bp in length
          *      55074  55173: gap of    100 bp
          *      55174  62527: contig of 7354 bp in length
          *      62528  62627: gap of    100 bp
          *      62628  66039: contig of 3412 bp in length
          *      66040  66139: gap of    100 bp
          *      66140  69939: contig of 3800 bp in length
          *      69940  70039: gap of    100 bp
          *      70040  78998: contig of 8959 bp in length
          *      78999  79098: gap of    100 bp
          *      79099  84516: contig of 5418 bp in length
          *      84517  84616: gap of    100 bp
          *      84617  93119: contig of 8503 bp in length
          *      93120  93219: gap of    100 bp
          *      93220  95763: contig of 2544 bp in length
          *      95764  95863: gap of    100 bp
          *      95864 109619: contig of 13756 bp in length
          *     109620 109719: gap of    100 bp
          *     109720 111843: contig of 2124 bp in length
          *     111844 111943: gap of    100 bp
          *     111944 114312: contig of 2369 bp in length
          *     114313 114412: gap of    100 bp
          *     114413 126023: contig of 11611 bp in length
          *     126024 126123: gap of    100 bp
          *     126124 130297: contig of 4174 bp in length
          *     130298 130397: gap of    100 bp
          *     130398 133226: contig of 2829 bp in length.
FEATURES             Location/Qualifiers
     source          1..133226
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-574F19"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..4495
                     /note="assembly_fragment:00744
                     fragment_chain:1"
     misc_feature    4596..17538
                     /note="assembly_fragment:00397
                     fragment_chain:1"
     misc_feature    17639..21198
                     /note="assembly_fragment:00155
                     fragment_chain:1"
     misc_feature    21299..24416
                     /note="assembly_fragment:00059
                     fragment_chain:2"
     misc_feature    24517..27345
                     /note="assembly_fragment:00653
                     fragment_chain:2"
     misc_feature    27946..35081
                     /note="assembly_fragment:00142
                     fragment_chain:3"
     misc_feature    35182..37225
                     /note="assembly_fragment:01217
                     fragment_chain:3"
     misc_feature    37326..39611
                     /note="assembly_fragment:00287
                     fragment_chain:4"
     misc_feature    39712..42865
                     /note="assembly_fragment:01572
                     fragment_chain:4"
     misc_feature    42966..49369
                     /note="assembly_fragment:00349
                     fragment_chain:5"
     misc_feature    49470..52200
                     /note="assembly_fragment:00572
                     fragment_chain:5"
     misc_feature    52301..55073
                     /note="assembly_fragment:00399
                     fragment_chain:6"
     misc_feature    55174..62527
                     /note="assembly_fragment:01761
                     fragment_chain:6"
     misc_feature    62628..66039
                     /note="assembly_fragment:00508
                     fragment_chain:7"
     misc_feature    66140..69939
                     /note="assembly_fragment:01732
                     fragment_chain:7"
     misc_feature    70040..78998
                     /note="assembly_fragment:01083
                     fragment_chain:8"
     misc_feature    79099..84516
                     /note="assembly_fragment:00253
                     fragment_chain:8"
     misc_feature    84617..93119
                     /note="assembly_fragment:01719
                     fragment_chain:9"
     misc_feature    93220..95763
                     /note="assembly_fragment:01419
                     fragment_chain:9"
     misc_feature    95864..109619
                     /note="assembly_fragment:00001"
```

```
        misc_feature        109720..111843
                            /note="assembly_fragment:00525"
        misc_feature        111944..114312
                            /note="assembly_fragment:00785"
        misc_feature        114413..126023
                            /note="assembly_fragment:01103"
        misc_feature        126124..130297
                            /note="assembly_fragment:01207"
        misc_feature        130398..133226
                            /note="assembly_fragment:01360"
BASE COUNT        38629 a    24563 c    25616 g    41987 t    2431 others
ORIGIN
```

```
                    Center: Wellcome Trust Sanger Institute
                    Center code: SC
                    Web site: http://www.sanger.ac.uk
                    Contact: humquery@sanger.ac.uk
                    ------------------------------------
FEATURES            Location/Qualifiers
        source      1..135820
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /clone="RP4-669K10"
                    /clone_lib="RPCI-4"
BASE COUNT    35410 a    32344 c    3048 g    35018 t
ORIGIN
```

Alignment Scores:
```
Pred. No.;              3.3e+03                    Length:              133226
Score:                  53.00                      Matches:             11
Percent Similarity:     75.00%                     Conservative:        1
Best Local Similarity:  68.75%                     Mismatches:          4
Query Match:            50.96%                     Indels:              0
DB:                     2                          Gaps:                0
```

US-09-833-017B-4 (1-21) x AL512364 (1-133226)

```
Qy      2 GlySerLeuThrPheProPheArgLeuPheAsnArgSerPheThrGln 17
          ||| ||||||||||||||||||||||||||::||  |||||||||
Db  46992 GGTGGGTGAAGCACGTTTGGGAGGCTTTTAACAAAGGTTCACACAG 47039
```

RESULT 15
AL513497/c

```
LOCUS       AL513497               135820 bp    DNA     linear   PRI 09-AUG-2002
DEFINITION  Human DNA sequence from clone RP4-669K10 on chromosome 1, complete
            sequence.
ACCESSION   AL513497
VERSION     AL513497.21  GI:22204196
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135820)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 11, 2002 this sequence version replaced gi:21530913.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-669K10 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            ------------ Genome Center
```

Alignment Scores:
```
Pred. No.:              3.37e+03                   Length:              135820
Score:                  53.00                      Matches:             11
Percent Similarity:     77.78%                     Conservative:        3
Best Local Similarity:  61.11%                     Mismatches:          4
Query Match:            50.96%                     Indels:              0
DB:                     9                          Gaps:                0
```

US-09-833-017B-4 (1-21) x AL513497 (1-135820)

```
Qy      1 SerGlySerLeuThrPheProPheArgLeuPheAsnArgSerPheThrGlnAla 18
          |||||||||||:|||||  |||||   |||::||  :||||  |||||||||||
Db  40215 AGCGGCTCAGTAAGCCTTTTTTTTTGAGATGGAGTCGTCGCTTCTTCACCCAGGCT 40162
```

Search completed: November 8, 2002, 22:57:36
Job time : 897.284 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2002, 20:29:27 ; Search time 104.687 Seconds
(without alignments)
451.748 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 104
Sequence: 1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
FgaPop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105348.4824/app_query.fasta_1.398
-DB=N.Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017_@CGN_1_1.125 @runat_05112002_105348.4824 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT - LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Length	% Match	DB Length	ID	Description
1		104	63	100.0	24	AAD32790	Streptococcus muta
2		104	63	100.0	24	AAD32884	Streptococcus muta
3		104	141	100.0	24	AAD32791	Streptococcus muta
4		104	141	100.0	24	AAD32898	Streptococcus muta
5		104	2557	100.0	24	AAD32800	Streptococcus muta
6		104	2557	100.0	24	AAD32893	Streptococcus muta
7	c	52	10475	50.0	9	AAN80616	Genomic clone enco
8	c	52	10475	50.0	9	AAN80643	Sequence of human
9	c	52	10476	50.0	24	ABL67064	Thyroid cancer rel
10	c	52	10476	50.0	24	ABL67518	Thyroid cancer rel
11		50	2017	48.1	22	AAS56980	C. trachomatis DNA
12		50	2932	48.1	23	ABL04050	Drosophila melanog
13		50	3596	48.1	23	ABL03962	Drosophila melanog
14	c	49	694	47.1	19	AAV27411	Streptococcus pneu
15	c	49	694	47.1	24	ABQ84879	S. pneumoniae SP10
16	c	49	1017	47.1	21	AAC41522	Arabidopsis thalia
17	c	49	1099	47.1	21	AAC32612	Arabidopsis thalia
18	c	49	1245	47.1	21	AAC35111	Arabidopsis thalia
19		49	2753	47.1	13	AAQ30801	pRi8 plasmid. Syn
20	c	49	2784	47.1	23	ABL08342	Drosophila melanog
21		49	3319	47.1	13	AAQ30802	Ri gene. Syntheti
22		49	3671	47.1	24	ABK84195	Human cDNA differe
23		49	3690	47.1	22	AAS31177	Human diagnostic a
24	c	49	4276	47.1	23	ABL11534	Drosophila melanog
25	c	49	6004	47.1	19	AAV52160	Streptococcus pneu
26		48.5	370	46.6	23	ABV37598	Human prostate exp
27		48	369	46.2	22	AAI98602	Human excretory re
28		48	369	46.2	22	AAI62998	Human kidney relat
29		48	885	46.2	24	ABQ46476	Oligonucleotide fo
30	c	48	885	46.2	24	ABQ46477	Oligonucleotide fo
31	c	48	1167	46.2	22	AAH52612	S. epidermidis ope
32	c	48	1317	46.2	24	ABN90839	Staphylococcus epi
33		48	3191	46.2	24	AAH54281	S. epidermidis gen
34		48	3285	46.2	22	AAH54084	S. epidermidis gen
35	c	48	3391	46.2	22	AAH54393	S. epidermidis gen
36		48	3682	46.2	23	ABL20364	Drosophila melanog
37		48	319608	46.2	21	AAH51601	Human chromosome 1
38		48	319608	46.2	24	AAS09301	Human schizophreni
39	c	47.5	565	45.7	24	ABN65805	Human cancer relat
40	c	47.5	3679	45.7	23	ABA93705	Human nucleic acid
41	c	47.5	3776	45.7	23	ABL22430	Drosophila melanog
42		47	870	45.2	24	ABN98796	Arabidopsis thalia
43		47	902	45.2	24	ABK30867	Plant dwarfing/stu
44		47	902	45.2	24	ABK30952	Plant dwarfing/stu
45		47	945	45.2	24	ABL49401	Sequence #3 used t

ALIGNMENTS

RESULT 1
AAD32790
ID   AAD32790 standard; DNA; 63 BP.
XX
AC   AAD32790;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans competence signal peptide (CSP) DNA.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..63

```
XX
DE   Streptococcus mutans competence signal peptide (CSP) DNA.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
KW   therapy; caries; endocarditis; microbial biofilm; infection; gene;
KW   vaccine; antibacterial; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..63
FT                   /tag= a
FT                   /product= "S. mutans competence signal peptide"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   CA2332733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/) LI Y H.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LAUP/) LAU P C.
XX
PI   Li YH, Cvitkovitch DG, Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB: AAE20584.
XX
PT   Novel compound that competitively inhibits binding of competence signal
PT   peptide to streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis -
XX
PS   Claim 8; Fig 2; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans CSP DNA.
XX
SQ   Sequence 63 BP; 20 A; 12 C; 13 G; 18 T; 0 other;

Alignment Scores:
Pred. No.:                2.65e-10        Length:          63
Score:                    104.00          Matches:         21
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       24              Gaps:            0

US-09-833-017B-4 (1-21) x AAD32884 (1-63)
```

```
FT                   /tag= a
FT                   /product= "S. mutans competence signal peptide (CSP)"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCY, Cvitkovitch DG, Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB: AAE20493.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis -
XX
PS   Claim 8; Fig 2; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. Vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans CSP DNA.
XX
SQ   Sequence 63 BP; 20 A; 12 C; 13 G; 18 T; 0 other;

Alignment Scores:
Pred. No.:                2.65e-10        Length:          63
Score:                    104.00          Matches:         21
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       24              Gaps:            0

US-09-833-017B-4 (1-21) x AAD32790 (1-63)

Qy    1  SerGlySerLeuThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly  20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  AGCGGAAGCCTATCAACATTTTCCGCGTGTTTAACAGAAGTTTTACACAAGCTTTGGGA  60

Qy   21  Lys  21
            |||
Db   61  AAA  63

RESULT 2
AAD32884
ID   AAD32884 standard; DNA; 63 BP.
XX
AC   AAD32884;
XX
DT   01-JUL-2002  (first entry)
```

Qy   1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 60

Qy   21 Lys 21
            |||
Db   61 AAA 63

RESULT 3
AAD32791
ID   AAD32791 standard; DNA; 141 BP.
XX
AC   AAD32791;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans comC gene.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; gene; comC gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..141
FT   /*tag= a
FT   /product= "S. mutans ComC protein"
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCY, Cvitkovitch DG, Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB; AAE20494.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis
XX
PS   Disclosure; Fig 2; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. Vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans comC CSP precursor gene.
XX
SQ   Sequence 141 BP; 58 A; 19 C; 25 G; 39 T; 0 other;

Alignment Scores:

Pred. No.:   7.13e-10   Length: 141
Score: 104.00   Matches: 21
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 24   Gaps: 0

US-09-833-017B-4 (1-21) x AAD32791 (1-141)

Qy   1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 135

Qy   21 Lys 21
            |||
Db   136 AAA 138

RESULT 4
AAD32898
ID   AAD32898 standard; DNA; 141 BP.
XX
AC   AAD32898;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans ComC gene.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
KW   therapy; caries; endocarditis; microbial.biofilm; infection; gene;
KW   vaccine; antibacterial; ComC gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..141
FT   /*tag= a
FT   /product= "S. mutans ComC protein"
XX
PN   CA2332733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/) LI Y H.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LAUP/) LAU P C.
XX
PI   Li YH, Cvitkovitch DG, Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB; AAE20594.
XX
PT   Novel compound that competitively inhibits binding of competence signal
PT   peptide to Streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis
XX
PS   Disclosure; Fig 2; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.

```
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans ComC gene encoding a
CC   CSP precursor protein.
XX
SQ   Sequence 141 BP; 58 A; 19 C; 25 G; 39 T; 0 other;

Alignment Scores:
Pred. No.:              7.13e-10        Length:              141
Score:                  104.00          Matches:             21
Percent Similarity:     100.00%         Conservative:        0
Best Local Similarity:  100.00%         Mismatches:          0
Query Match:            24              Indels:              0
DB:                                     Gaps:                0

US-09-833-017B-4 (1-21) x AAD32898 (1-141)

Qy     1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    76 AGCGGAAGCCTATCAACACATTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGGA 135

Qy    21 Lys 21
         |||
Db   136 AAA 138

RESULT 5
AAD32800
ID   AAD32800 standard; DNA; 2557 BP.
XX
AC   AAD32800;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus mutans comCDE gene local region.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; therapy; caries;
KW   endocarditis; immunopurification; antibacterial; antiinflammatory;
KW   genetic competence assay; vaccine; comCDE gene; gene; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key          Location/Qualifiers
FT   CDS          complement (25..168)
FT                /*tag= a
FT                /product= "Protein #1 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          complement (452..499)
FT                /*tag= b
FT                /product= "Peptide #1 encoded by S. mutans comCDE gene
FT                local region"
FT                /note="CDS does not include start and stop codon"
FT                /partial
FT   CDS          648..758
FT                /*tag= c
FT                /product= "Peptide #2 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          953..1081
FT                /*tag= d
FT                /product= "Protein #2 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          complement (1366..1449)
FT                /*tag= e
FT                /product= "Peptide #3 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          1855..1959
FT                /*tag= f
FT                /product= "Peptide #4 encoded by S. mutans comCDE gene

FT                local region"
FT   CDS          complement (1896..2072)
FT                /*tag= g
FT                /product= "Protein #3 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          2182..2424
FT                /*tag= h
FT                /product= "Protein #4 encoded by S. mutans comCDE gene
FT                local region"
FT   CDS          2384..2488
FT                /*tag= i
FT                /product= "Peptide #5 encoded by S. mutans comCDE gene
FT                local region"
XX
PN   CA2302861-A1.
XX
PD   10-OCT-2001.
XX
PF   10-APR-2000; 2000CA-2302861.
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LAUP/) LAU P C Y.
PA   (CVIT/) CVITKOVITCH D G.
PA   (LIYH/) LI Y H.
XX
PI   Lau PCY, Cvitkovitch DG,   Li YH;
XX
DR   WPI; 2002-242151/30.
DR   P-PSDB; AAE20607, AAE20608, AAE20609, AAE20610, AAE20611, AAE20612,
DR   AAE20613, AAE20614, AAE20615.
XX
PT   Novel compound that inhibits binding of competence signal peptide of
PT   Streptococcus mutans to S. mutans histidine kinase, useful for treating
PT   or preventing caries or endocarditis -
XX
PS   Disclosure; Fig 9; 54pp; English.
XX
CC   The invention relates to compound that competitively inhibits binding
CC   of competence signal peptide (CSP) to Streptococcus mutans histidine
CC   kinase (HK). Compounds of the invention are useful for treating or
CC   prophylaxis of caries or endocarditis. Vector comprising nucleic acid
CC   encoding S. mutans CSP is useful for producing recombinant host cell
CC   capable of expressing it. The recombinant host cell produced by the
CC   method is useful for expressing peptide in culture. S. mutans CSP DNA
CC   is useful for identifying nucleic acid molecules encoding CSP activated
CC   peptide. It is also useful as probes and in assays to identify
CC   antagonists or inhibitors of the peptides produced by the nucleic acid
CC   molecules. It is also useful for preparing vaccines for preventing or
CC   treating the above mentioned conditions. Antibodies against CSP activity
CC   are also useful for preventing caries. The antibodies are also useful
CC   for screening organisms or tissues containing CSP peptide or CSP-like
CC   peptide and for immunopurifying the peptides. The CSP nucleic acid
CC   molecules are useful in assays for genetic competence. The present
CC   sequence is S. mutans comCDE gene local region.
XX
SQ   Sequence 2557 BP; 891 A; 415 C; 426 G; 825 T; 0 other;

Alignment Scores:
Pred. No.:              2.5e-08         Length:              2557
Score:                  104.00          Matches:             21
Percent Similarity:     100.00%         Conservative:        0
Best Local Similarity:  100.00%         Mismatches:          0
Query Match:            24              Indels:              0
DB:                                     Gaps:                0

US-09-833-017B-4 (1-21) x AAD32800 (1-2557)

Qy     1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176 AGCGGAAGCCTATCAACACATTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGGA 235

Qy    21 Lys 21
```

```
Db        236 AAA 238
          | | |
          236 AAA 238

RESULT 6
AAD32893
ID   AAD32893 standard; DNA; 2557 BP.
XX
AC   AAD32893;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus mutans comCDE gene local region.
XX
KW   Competence signal peptide; CSP; histidine kinase; HK; prophylaxis;
KW   therapy; caries; endocarditis; microbial biofilm; infection; gene;
KW   vaccine; antibacterial; comCDE gene local region; ds.
XX
OS   Streptococcus mutans.
XX
FH   Key            Location/Qualifiers
FT   CDS            complement (25..168)
FT                  /*tag= a
FT                  /product= "Protein #1 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  101..241
FT                  /*tag= b
FT                  /product= "S. mutans ComC protein"
FT   CDS            complement (383..1708)
FT                  /*tag= c
FT                  /product= "S. mutans ComD protein"
FT   CDS            complement (452..499)
FT                  /*tag= d
FT                  /product= "Peptide #1 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  /note= "CDS does not include start and stop codon"
FT                  /partial
FT                  648..758
FT                  /*tag= e
FT                  /product= "Peptide #2 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  953..1081
FT                  /*tag= f
FT   CDS            /product= "Protein #2 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  complement (1366..1449)
FT                  /*tag= g
FT                  /product= "Peptide #3 encoded by S. mutans ComCDE
FT                  gene local region"
FT   CDS            complement (1705..2457)
FT                  /*tag= h
FT                  /product= "S. mutans ComE protein"
FT                  1855..1959
FT                  /*tag= i
FT                  /product= "Peptide #4 encoded by S. mutans ComCDE
FT                  gene local region"
FT   CDS            complement (1896..2072)
FT                  /*tag= j
FT                  /product= "Protein #3 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  2182..2424
FT                  /*tag= k
FT                  /product= "Protein #4 encoded by S. mutans ComCDE
FT                  gene local region"
FT                  2384..2488
FT                  /*tag= l
FT   CDS            /product= "Peptide #5 encoded by S. mutans ComCDE
FT                  gene local region"
XX
PN   CA232733-A1.
XX
PD   10-OCT-2001.
XX
PF   20-FEB-2001; 2001CA-2332733.
```

```
XX
PR   10-APR-2000; 2000CA-2302861.
XX
PA   (LIYH/)  LI Y H.
PA   (CVIT/)  CVITKOVITCH D G.
PA   (LAUP/)  LAU P C.
XX
PI   Li YH,  Cvitkovitch DG,  Lau PC;
XX
DR   WPI; 2002-242173/30.
DR   P-PSDB; AAE20617, AAE20618, AAE20619, AAE20620, AAE20621, AAE20622,
DR   AAE20623, AAE20624, AAE20625, AAE20594, AAE20585, AAE20586.
XX
PT   Novel compound that competitively inhibits binding of competence signal
PT   peptide to Streptococcus mutans histidine kinase, useful in treatment
PT   or prophylaxis of caries or endocarditis -
XX
PS   Disclosure; Fig 9; 82pp; English.
XX
CC   The invention relates to a compound that competitively inhibits the
CC   binding of competence signal peptide (CSP) to Streptococcus mutans
CC   histidine kinase (HK). Compounds of the invention are useful in
CC   medical treatment or prophylaxis of caries or endocarditis. They are
CC   useful for inhibiting or disrupting microbial biofilms involved in
CC   infections in man and animals and in biofouling of surfaces susceptible
CC   to microbial accumulation. They are also useful for treatment or
CC   prophylaxis of a disease, disorder or abnormal physical state caused
CC   by S. mutans. Pharmaceutical composition containing the compounds of
CC   the invention is useful for treating diseases caused by streptococcal
CC   infections. Polynucleotides encoding S. mutans CSP are useful as probes
CC   or in assays to identify antagonists or inhibitors of CSP peptides.
CC   They are also used for preparing vaccines. Polypeptides of the invention
CC   are useful for preparing antibodies, for vitro analysis of HK, CSP or
CC   RR activity or structure, and in assays for the identification and
CC   developments of compounds to inhibit and/or enhance polypeptide or
CC   peptide function directly. Antibodies of the invention are useful for
CC   providing protection against caries, to screen organisms or tissues
CC   containing CSP peptide or CSP-like peptides, for immuno-purification of
CC   CSP or CSP-like peptides from crude extracts, and to detect CSP or a
CC   similar peptide. The present sequence is S. mutans comCDE gene local
CC   region.
XX
SQ   Sequence 2557 BP; 891 A; 415 C; 426 G; 825 T; 0 other;

Alignment Scores:
Pred. No.:               2.5e-08
Score:                   104.00          Length:                 2557
Percent Similarity:      100.00%         Matches:                21
Best Local Similarity:   100.00%         Conservative:           0
Query Match:             100.00%         Mismatches:             0
DB:                      24              Indels:                 0
                                         Gaps:                   0

us-09-833-017B-4 (1-21) x AAD32893 (1-2557)

Qy    1 SerGlySerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
         | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  176 AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGGA 235

Qy   21 Lys 21
         | | |
Db  236 AAA 238

RESULT 7
AAN80616/c
ID   AAN80616 standard; DNA; 10475 BP.
XX
AC   AAN80616;
XX
DT   03-OCT-2002  (updated)
DT   21-NOV-1990  (first entry)
XX
DE   Genomic clone encoding SAP(Phe).
XX
```

```
KW   SAP(Val); SAP(Phe); hyaline membrane disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   exon           1054..1120
FT                  /*tag= a
FT                  /label=exon I
FT   exon           1431..1558
FT                  /*tag= b
FT                  /label=exon II
FT   exon           2060..2131
FT                  /*tag= c
FT                  /label=exon III
FT   exon           2495..2620
FT                  /*tag= d
FT                  /label=exon IV
FT   exon           3447..3635
FT                  /*tag= e
FT                  /label=exon V
FT   exon           3863..3952
FT                  /*tag= f
FT                  /label=exon VI
FT   exon           5386..5568
FT                  /*tag= g
FT                  /label=exon VII
FT   exon           5742..5887
FT                  /*tag= h
FT   exon           7122..7202
FT                  /label=exon VIII
FT                  /*tag= i
FT   exon           7697..7756
FT                  /label=exon IX
FT                  /*tag= j
FT                  /label=exon X
XX
PN   WO8804324-A.
XX
PD   16-JUN-1988.
XX
PF   03-DEC-1987;   87WO-US03180.
XX
PR   03-DEC-1987;   87WO-US03180.
PR   03-DEC-1987;   87US-0939206.
XX
PA   (ABBO ) ABBOTT LABORATORIES.
XX
PI   Whitsett JA,  Fox JL,  Pilot-Matias TJ,  Meuth JL;
XX
DR   WPI; 1988-175472/25.
DR   P-PSDB; AAP82934.
XX
PT   Pulmonary hydrophobic surfactant-associated proteins - useful for
PT   normalising pulmonary surface tension.
XX
PS   Disclosure; ; PP; English.
XX
CC   This genomic clone encodes the pulmonary hydrophobic surfactant-
CC   associated protein SAP(Phe). Following processing to remove the
CC   introns the prod. is obtd. in a high yield. When SAP(Phe) is
CC   combined with a lipid it can be used to reduce or maintain
CC   normal pulmonary surface tension in the alveoli of animals
CC   (esp. humans) and therefore can be used in the treatment of
CC   hyaline membrane disease in premature infants.
CC   May also be used to deliver substances to respiratory epithelial cells.
CC   See also AAP80572-96, AAN80613, AAN80615, AAN80617 and AAN82412-13.
CC   (Updated on 03-OCT-2002 to add missing OS field.)
XX
SQ   Sequence 10475 BP; 2416 A; 3063 C; 2940 G; 2056 T; 0 other;

Alignment Scores:
Pred. No.:   215               Length:   10475
Score:       52.00             Matches:  10
```

```
Percent Similarity:      76.47%        Conservative:  3
Best Local Similarity:   58.82%        Mismatches:    4
Query Match:             50.00%        Indels:        0
DB:                                    Gaps:          0

US-09-833-017B-4 (1-21) x AAN80616 (1-10475)

Qy    1  SerGlySerLeuSerThrPhePheArgLeuAsnArgSerPheThrGln  17
         |||||||||:::  |||:::|||   |||:::||||||||
Db  543  TCTGGTTCATGGACACATCGTTTGCATTAGTCAGAAGAGCATTACACAA  493

RESULT 8
AAN80643/C
ID   AAN80643 standard; DNA; 10475 BP.
XX
AC   AAN80643;
XX
DT   03-FEB-1991  (first entry)
XX
DE   Sequence of human pulmonary hydrophobic surfactant-associated protein
DE   (SAP) (Phe) genomic clone.
XX
KW   Human pulmonary hydrophobic surfactant-associated protein (SAP) (Phe);
KW   hyaline membrane disease (HMD); therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   exon           1054..1119
FT                  /*tag= a
FT                  /note="AAP80193"
FT   exon           1433..1558
FT                  /codon_start=1433
FT                  /*tag= b
FT                  /note="AAP80194"
FT                  2059..2130
FT                  /*tag= c
FT                  /note="AAN80195"
FT   exon           2494..2619
FT                  /*tag= d
FT                  /note="AAP81070"
FT   exon           3445..3633
FT                  /*tag= e
FT                  /note="AAP81074"
FT                  3861..3950
FT                  /*tag= f
FT                  /note="AAP81075"
FT   exon           5384..5567
FT                  /*tag= g
FT                  /note="AAP81076"
FT   exon           5742..5885
FT                  /*tag= h
FT                  /codon_start=5742
FT                  /note="AAP81077"
FT   exon           7120..7200
FT                  /*tag= i
FT                  /note="AAP81078"
FT   exon           7695..7754
FT                  /*tag= j
FT                  /note="AAP81079"
XX
PN   WO8803170-A.
XX
PD   05-MAY-1988.
XX
PF   02-OCT-1987;   87WO-US02536.
XX
PR   01-OCT-1987;   87US-0101660.
PR   08-DEC-1986;   86US-0939206.
PR   10-JUN-1987;   87US-0060719.
XX
PA   (WHIT/) WHITSETT JA.
XX
```

```
PI   Whitsett JA, Fox JL, Pilot-Matias TJ, Meuth JL, Sarin VK;
XX
DR   WPI; 1988-133244/19.
DR   P-PSDB: AAP80193, AAP80194, AAP80195, 081070, AAP81074, AAP81075,
DR   AAP81076, AAP81077, AAP81078, AAP81079.
XX
PT   Pulmonary hydrophobic surfactant-associated proteins -
PT   used with lipid(s) to treat and prevent hyaline membrane disease
PT   and similar syndromes
XX
PS   Example; Fig 8B-1 - 8B-24; 144pp; English.
XX
CC   The sequence is derived from human embryonic kidney cell genomic
CC   library. SAP (Val) and SAP (Phe), when combined with lipids, have
CC   significant pulmonary biophysical surfactant activity that may be
CC   utilised to treat and prevent hyaline membrane disease (HMD) and
CC   other syndromes associated with lack or insufficient amts. of natural
CC   pulmonary surfactant material. Antibodies and antisera may also be
CC   made which are directed against SAP (Val) or SAP (Phe).
CC   SAP (Val) and SAP (Phe) in body fluids may be assayed using the compsns.
XX
SQ   Sequence 10475 BP; 2415 A; 3060 C; 2942 G; 2056 T; 2 other;

Alignment Scores:
Pred. No.:                215                Length:            10475
Score:                    52.00              Matches:           10
Percent Similarity:       76.47%             Conservative:      3
Best Local Similarity:    58.82%             Mismatches:        4
Query Match:              50.00%             Indels:            0
DB:                                          Gaps:              0

US-09-833-017B-4 (1-21) x AAN80643 (1-10475)

Qy     1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
          |||||||||::: ||||:::||| ||||     |||::||||||||||||||
Db   543  TCTGGTTCTATGGACACATGGTTTGCATTAGTCAGAAGAGCATTTACACAA 493

RESULT 9
ABL67064/C
ID   ABL67064 standard; DNA; 10476 BP.
XX
AC   ABL67064;
XX
DT   15-MAY-2002 (first entry)
XX
DE   Thyroid cancer related gene sequence SEQ ID NO:5401.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO20194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   20-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene in expression of a signature gene set -
XX
PS   Claim 1; SEQ ID 5401; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 10476 BP; 2416 A; 3059 C; 2944 G; 2057 T; 0 other;

Alignment Scores:
Pred. No.:                215                Length:            10476
Score:                    52.00              Matches:           10
Percent Similarity:       76.47%             Conservative:      3
Best Local Similarity:    58.82%             Mismatches:        4
Query Match:              50.00%             Indels:            0
DB:                                          Gaps:              0

US-09-833-017B-4 (1-21) x ABL67064 (1-10476)
```

```
Qy      1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
          |||||||||:::|||:|||   |||   |||:::|||||||||
Db    543 TCTGGTTCTATGGACACATGGTTGCATTAGTCAGAAGAGCATTTACACAA 493

RESULT 10
ABL67518/c
ID   ABL67518 standard; DNA; 10476 BP.
XX
AC   ABL67518;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Thyroid cancer related gene sequence SEQ ID NO:5855.
XX
KW   Human; cancer: colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   20-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-18264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 5855; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 10476 BP; 2416 A; 3059 C; 2944 G; 2057 T; 0 other;

Alignment Scores:
Pred. No.:                              Length:        10476
Score:              215                 Matches:       10
Percent Similarity:  52.00              Conservative:  3
Best Local Similarity: 76.47%           Mismatches:    4
Query Match:        58.82%              Indels:        0
DB:                 50.00%              Gaps:          0
                    24

US-09-833-017B-4 (1-21) x ABL67518 (1-10476)

Qy      1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
          |||||||||:::|||:|||   |||   |||:::|||||||||
Db    543 TCTGGTTCTATGGACACATGGTTGCATTAGTCAGAAGAGCATTTACACAA 493

RESULT 11
AAS56980
ID   AAS56980 standard; DNA; 2017 BP.
XX
AC   AAS56980;
XX
DT   16-JAN-2002  (first entry)
XX
DE   C. trachomatis DNA clone E1-A5-53 encoding partial dnaK,grpE and CT166.
XX
KW   Chlamydia; ds; sexually transmitted disease; PID; antibacterial;
KW   pelvic inflammatory disease; antigen; trachoma; gynecological;
KW   acute respiratory tract infection; atherosclerosis; male infertility;
KW   coronary heart disease.
XX
OS   Chlamydia trachomatis.
XX
PN   WO200181379-A2.
XX
PD   01-NOV-2001.
XX
PF   23-APR-2001; 2001WO-US13081.
XX
PR   21-APR-2000; 2000US-198853P.
PR   20-JUL-2000; 2000US-219752P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Bhatia A, Probst P, Stromberg EJ;
```

XX
DR   WPI; 2001-616771/71.
XX
PT   New polynucleotide for treating Chlamydia infections encodes a
PT   polynucleotides containing an immunogenic portion of a Chlamydia
PT   antigen -
XX
PS   Claim 1; Page 116-117; 208pp; English.
XX
CC   The invention relates to isolated polynucleotide encoding at least
CC   a partial Chlamydia protein which is an antigenic fragment, or the
CC   complements, fragments, homologues and variants, and antibodies
CC   raised against the antigenic proteins (or fragments). The nucleic
CC   acids, proteins and antibodies are used to diagnose and treat Chlamydia
CC   infections (e.g. a sexually transmitted disease, pelvic inflammatory
CC   disease (PID), acute respiratory tract infection, trachoma,
CC   atherosclerosis and coronary heart disease) in a patient, and in
CC   the treatment of male infertility. The compounds of the invention are
CC   also useful for detecting the presence of Chlamydia in a patient, and
CC   stimulating and/or expanding T cells specific for a Chlamydia protein.
CC   The present sequence encodes a Chlamydia antigen.
XX
SQ   Sequence 2017 BP; 490 A; 443 C; 389 G; 695 T; 0 other;

Alignment Scores:
Pred. No.:   64.3   Length: 2017
Score: 50.00   Matches: 10
Percent Similarity: 78.57%   Conservative: 1
Best Local Similarity: 71.43%   Mismatches: 3
Query Match: 48.08%   Indels: 0
DB: 22   Gaps: 0

US-09-833-017B-4 (1-21) x AAS56980 (1-2017)

Qy   3 SerLeuSerThrPheAlaArgLeuPheAsnArgSerPheThr 16
          ::||| |||||||| |||||||||||||
Db   353 GCACTAAATGACCTTTTCAGTTTATTTAATAGGTCTTTAACA 394

RESULT 12
ABL04050
ID   ABL04050 standard; cDNA; 2932 BP.
XX
AC   ABL04050;
XX
DT   26-MAR-2002   (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6632.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59947.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX

PS   Claim 1; SEQ ID NO 6632; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL16175), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2932 BP; 859 A; 603 C; 591 G; 879 T; 0 other;

Alignment Scores:
Pred. No.:   102   Length: 2932
Score: 50.00   Matches: 10
Percent Similarity: 80.00%   Conservative: 2
Best Local Similarity: 66.67%   Mismatches: 3
Query Match: 48.08%   Indels: 0
DB: 23   Gaps: 0

US-09-833-017B-4 (1-21) x ABL04050 (1-2932)

Qy   1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 15
          ||||||||||||||| :::|| ||||||::|||
Db   2127 TCAGGATCGTTGAGTTTATATTTCTATTTATTTTCTCGGGATTTT 2171

RESULT 13
ABL03962
ID   ABL03962 standard; cDNA; 3596 BP.
XX
AC   ABL03962;
XX
DT   26-MAR-2002   (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 6368.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59859.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 6368; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

Left column:

```
CC    sequences (ABL01840-ABL16175) and the encoded proteins
CC    (ABB7737-ABB72072).
CC    The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 3596 BP; 898 A; 897 C; 809 G; 992 T; 0 other;

Alignment Scores:
Pred. No.:              131           Length:           3596
Score:                  50.00         Matches:          10
Percent Similarity:     80.00%        Conservative:     2
Best Local Similarity:  66.67%        Mismatches:       3
Query Match:            48.08%        Indels:           0
DB:                     23            Gaps:             0

US-09-833-017B-4 (1-21) x ABL03962 (1-3596)

Qy     1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPhe 15
           ||||||||||||||    :::||   |||    |||
Db   831 TCAGGATCGTTGAGTTTATATTTCTATTTATTTTCTCGGGATTT 875

RESULT 14
AAV27411/c
ID   AAV27411 standard; DNA; 694 BP.
XX
AC   AAV27411;
XX
DT   02-OCT-1998 (first entry)
XX
DE   Streptococcus pneumoniae SP0100 nucleotide.
XX
KW   Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW   detection; pneumonia; otitis media; meningitis; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key            Location/Qualifiers
FT   CDS            2..694
FT                  /*tag= a
FT                  /product= "SP0100"
FT                  /transl_except= (pos:590..592,aa:Xaa)
FT                  /note= "no stop codon given; Xaa is unspecified"
XX
PN   WO9818930-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;     97WO-US19422.
XX
PR   31-OCT-1996;     96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR   WPI; 1998-272224/24.
DR   P-PSDB; AAW61225.
XX
PT   Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT   pneumoniae - or their epitope-containing fragments, useful in
PT   protective or therapeutic vaccines, and for diagnosis .
XX
PS   Claim 1; Page 84; 118pp; English.
XX
CC   The present sequence encodes a protein from Streptococcus pneumoniae.
CC   The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC   can be useful in vaccines for inducing protective antibodies against
CC   Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC   pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC   are used to detect Streptococcus pneumoniae infection (by usual hybridisation or
CC   amplification methods), also for isolating Streptococcus pneumoniae genes or their
CC   allelic variants. The protein can be used similarly to detect specific
```

Right column:

```
CC   antibodies in standard immunoassays, especially for diagnosing or
CC   monitoring infections. Antibodies which bind the protein are used to
CC   detect corresponding antigens, to purify the protein and for passive
CC   immunisation (optionally coupled to a toxin). Vaccines are administered,
CC   e.g. by injection, orally or through the skin, typically at 0.01-1000
CC   (especially 10-300) mu g/ml per dose.
XX
SQ   Sequence 694 BP; 271 A; 77 C; 111 G; 234 T; 1 other;

Alignment Scores:
Pred. No.:              26.1          Length:           694
Score:                  49.00         Matches:          10
Percent Similarity:     70.59%        Conservative:     2
Best Local Similarity:  58.82%        Mismatches:       5
Query Match:            47.12%        Indels:           0
DB:                     19            Gaps:             0

US-09-833-017B-4 (1-21) x AAV27411 (1-694)

Qy     5 SerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
           ||||||||      ||| :||:|||||||| :::|||||||||
Db   210 TCTACCTTTTCTCCGTTTTTCACTAAAAGTTTCACGGCGTTCTCTAGGAAAA 160

RESULT 15
ABQ84879/c
ID   ABQ84879 standard; DNA; 694 BP.
XX
AC   ABQ84879;
XX
DT   04-SEP-2002 (first entry)
XX
DE   S. pneumoniae SP100 nucleotide sequence SEQ ID NO:175.
XX
KW   Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW   antibacterial; Streptococcal infection; detection; gene; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   US2002061545-A1.
XX
PD   23-MAY-2002.
XX
PF   22-JAN-2001; 2001US-0765272.
XX
PR   30-OCT-1997;     97US-0961083.
XX
PA   (CHOL/) CHOI G H.
PA   (KUNS/) KUNSCH C A.
PA   (BARA/) BARASH S C.
PA   (DILL/) DILLON P J.
PA   (DOUG/) DOUGHERTY B.
PA   (FANN/) FANNON M R.
PA   (ROSE/) ROSEN C A.
XX
PI   Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI   Rosen CA;
XX
DR   WPI; 2002-479261/51.
DR   P-PSDB; ABP54644.
XX
PT   New Streptococcus pneumoniae antigens, useful for detecting
PT   Streptococcus and for preventing or attenuating disease caused by
PT   Streptococcus infection -
XX
PS   Claim 1; Page 44; 70pp; English.
XX
CC   ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC   Streptococcus pneumoniae protein antigens given in ABP54557 to ABP54669.
CC   The S. pneumoniae antigens have antibacterial activity and can be
CC   used in vaccines. The S. pneumoniae antigens can also be used to
CC   prevent or attenuate a Streptococcal infection in an animal. The
CC   polynucleotides encoding the S. pneumoniae antigens can be used to
CC   detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
```

CC   primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC   which are used in an example from the present invention.
XX
SQ   Sequence 694 BP; 271 A; 77 C; 111 G; 234 T; 1 other;

Alignment Scores:
Pred. No.:   26.1   Length: 694
Score: 49.00   Matches: 10
Percent Similarity: 70.59%   Conservative: 2
Best Local Similarity: 58.82%   Mismatches: 5
Query Match: 47.12%   Indels: 0
DB: 24   Gaps: 0

US-09-833-017B-4 (1-21) x ABQ84879 (1-694)

Qy   5 SerThrPhePheArgPheLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
          |||||||||| ||| :::||||||||| ::::|||||||||
Db   210 TCTACCTTTTCTCCGTTTTCACTAAAAGTTTCACGGCTTCTCTAGGAAAA   160

Search completed: November 8, 2002, 22:11:00
Job time : 108.687 secs

```
                    GenCore version 5.1.3
             Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      November  8, 2002, 22:05:37 ; Search time 21.9403 Seconds
                                     (without alignments)
                                     293.533 Million cell updates/sec

Title:           US-09-833-017B-4
Perfect score:   104
Sequence:        1 SGSLSTFPRLFNRSFTQALGK 21

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:       882724

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105349_4858/app_query.fasta_1.398
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017 @CGN_1_17_@runat_05112002_105349_4858 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    Issued_Patents_NA:*
              1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
              2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
              3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
              4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
              5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
              6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	49	47.1	694	3	US-08-961-083-175	Sequence 175, App
c 2	49	47.1	2756	1	US-08-187-793-1	Sequence 1, Appli
c 3	49	47.1	3318	1	US-08-187-793-3	Sequence 3, Appli
c 4	49	47.1	6004	4	US-08-961-527-27	Sequence 27, Appl
c 5	48	46.2	1317	4	US-09-134-001C-302	Sequence 302, App
c 6	46.5	44.7	2682	1	US-07-971-819A-1	Sequence 1, Appli
c 7	46.5	44.7	2682	1	US-07-977-434-3	Sequence 3, Appli
c 8	46.5	44.7	2682	1	US-08-475-231-1	Sequence 1, Appli
c 9	46.5	44.7	2682	1	US-08-458-819-3	Sequence 3, Appli
c 10	46.5	44.7	2682	4	US-09-105-697-9	Sequence 9, Appli
c 11	46.5	44.7	2682	5	PCT-US91-07035-3	Sequence 3, Appli
c 12	45	43.3	1651	4	US-09-484-970B-17	Sequence 17, Appl
c 13	45	43.3	8411	4	US-08-961-527-16	Sequence 16, Appl
c 14	45	43.3	65042	4	US-09-784-316-3	Sequence 3, Appl1
c 15	44	42.3	1179	4	US-08-745-995A-22	Sequence 22, Appl
16	44	42.3	1179	4	US-08-745-995A-23	Sequence 23, Appl
c 17	44	42.3	1191	4	US-08-745-995A-4	Sequence 4, Appli
18	44	42.3	1191	4	US-08-745-995A-5	Sequence 5, Appl1
c 19	44	42.3	1191	4	US-08-745-995A-34	Sequence 34, Appl
20	44	42.3	1191	4	US-08-745-995A-35	Sequence 35, Appl
c 21	44	42.3	1194	4	US-08-745-995A-28	Sequence 28, Appl
22	44	42.3	1194	4	US-08-745-995A-29	Sequence 29, Appl
23	44	42.3	1197	4	US-08-745-995A-10	Sequence 10, Appl
c 24	44	42.3	1197	4	US-08-745-995A-11	Sequence 11, Appl
c 25	44	42.3	1260	4	US-08-745-995A-16	Sequence 16, Appl
26	44	42.3	1260	4	US-08-745-995A-17	Sequence 17, Appl
27	44	42.3	1358	4	US-08-745-995A-7	Sequence 7, Appli
c 28	44	42.3	1358	4	US-08-745-995A-9	Sequence 9, Appli
c 29	44	42.3	1414	4	US-08-745-995A-19	Sequence 19, Appl
30	44	42.3	1414	4	US-08-745-995A-21	Sequence 21, Appl
c 31	44	42.3	1454	4	US-08-745-995A-31	Sequence 31, Appl
32	44	42.3	1454	4	US-08-745-995A-33	Sequence 33, Appl
c 33	44	42.3	1492	4	US-08-745-995A-25	Sequence 25, Appl
34	44	42.3	1492	4	US-08-745-995A-27	Sequence 27, Appl
c 35	44	42.3	1584	4	US-08-745-995A-1	Sequence 1, Appli
36	44	42.3	1584	4	US-08-745-995A-3	Sequence 3, Appli
c 37	44	42.3	1838	4	US-08-745-995A-13	Sequence 13, Appl
38	44	42.3	1838	4	US-08-745-995A-15	Sequence 15, Appl
39	43	41.3	510	4	US-09-134-001C-1753	Sequence 1753, Ap
40	43	41.3	777	4	US-08-809-326A-7	Sequence 7, Appli
c 41	43	41.3	813	4	US-08-809-326A-4	Sequence 4, Appli
c 42	43	41.3	1000	4	US-09-641-638-647	Sequence 647, App
43	43	41.3	1048	4	US-08-809-326A-9	Sequence 9, Appli
c 44	43	41.3	1296	4	US-08-809-326A-18	Sequence 18, Appl
c 45	43	41.3	1464	4	US-08-809-326A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-961-083-175/c
; Sequence 175, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
;    APPLICANT: Choi et. al.
;    TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;    NUMBER OF SEQUENCES: 452
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Human Genome Sciences, Inc.
;       STREET: 9410 Key West Avenue
;       CITY: Rockville
;       STATE: Maryland
;       COUNTRY: USA
;       ZIP: 20850
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER: HP Vectra 486/33
;       OPERATING SYSTEM: MSDOS version 6.2
;       SOFTWARE: ASCII Text
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/961,083
;       FILING DATE:
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Brookes, A, Anders
;       REGISTRATION NUMBER: 36,373
;       REFERENCE/DOCKET NUMBER: PB340P2
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (301) 309-8504
;       TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 175:
;    SEQUENCE CHARACTERISTICS:
```

```
      ;      LENGTH:   694 base pairs
      ;      TYPE:  nucleic acid
      ;      STRANDEDNESS:  double
      ;      TOPOLOGY:  linear
US-08-961-083-175

Alignment Scores:
Pred. No.:               5.46           Length:           694
Score:                  49.00           Matches:           10
Percent Similarity:     70.59%          Conservative:       2
Best Local Similarity:  58.82%          Mismatches:         5
Query Match:            47.12%          Indels:             0
DB:                         3           Gaps:               0

US-09-833-017B-4 (1-21) x US-08-961-083-175 (1-694)

Qy     5  SerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys  21
           ||||||||||         |||  ::||||||||||  ::||||||||||
Db   210  TCTACCTTTTCTCCGGTTTTCACTAAAAGTTTCACCGCTTCTCTAGGAAAA  160

RESULT 2
US-08-187-793-1
; Sequence 1, Application US/08187793
; Patent No. 5614371
;  GENERAL INFORMATION:
;    APPLICANT:  POSNER, JEROME B.
;    APPLICANT:  DARNELL, ROBERT B.
;    APPLICANT:  FURNEAUX, HENRY M.
;    TITLE OF INVENTION:  ANTIGEN RECOGNIZED BY ANTIBODIES ASSOCIATED
;    TITLE OF INVENTION:  WITH PARANEOPLASTIC OPSOCLONUS AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER & DUNHAM
;      STREET:  30 ROCKEFELLER PLAZA
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,793
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 691,559
;      FILING DATE:  25-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  39227
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  (212) 422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2756 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1029
;      OTHER INFORMATION:
US-08-187-793-1
```

```
Alignment Scores:
Pred. No.:              29.4            Length:          2756
Score:                  49.00           Matches:           10
Percent Similarity:     91.67%          Conservative:       1
Best Local Similarity:  83.33%          Mismatches:         1
Query Match:            47.12%          Indels:             0
DB:                         1           Gaps:               0

US-09-833-017B-4 (1-21) x US-08-187-793-1 (1-2756)

Qy     4  LeuSerThrPhePheArgLeuPheAsnArgSerPhe  15
           |||||||||||   ||||||||||:::||||||||
Db  2197  TTATCTACCTTTGTCAGGCTTTTGACAGGTCATTT  2232

RESULT 3
US-08-187-793-3
; Sequence 3, Application US/08187793
; Patent No. 5614371
;  GENERAL INFORMATION:
;    APPLICANT:  POSNER, JEROME B.
;    APPLICANT:  DARNELL, ROBERT B.
;    APPLICANT:  FURNEAUX, HENRY M.
;    TITLE OF INVENTION:  ANTIGEN RECOGNIZED BY ANTIBODIES ASSOCIATED
;    TITLE OF INVENTION:  WITH PARANEOPLASTIC OPSOCLONUS AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER & DUNHAM
;      STREET:  30 ROCKEFELLER PLAZA
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,793
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 691,559
;      FILING DATE:  25-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  39227
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  (212) 422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1590
;      OTHER INFORMATION:
US-08-187-793-3

Alignment Scores:
Pred. No.:              36.9            Length:          3318
Score:                  49.00           Matches:           10
Percent Similarity:     91.67%          Conservative:       1
```

Best Local Similarity: 83.33%   Mismatches: 1
Query Match: 47.12%   Indels: 0
DB: Gaps: 0

US-09-833-017B-4 (1-21) x US-08-187-793-3 (1-3318)

Qy   4 LeuSerThrPheArgLeuPheAsnArgSerPhe 15
        ||||||||||||||::::||||||||||||||||
Db 2758 TTATCTACCTTTGTCAGGCTTTTTGACAGGTCATTT 2793

RESULT 4
US-08-961-527-27
; Sequence 27, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; ; APPLICANT: Charles Kunsch
; ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; ; NUMBER OF SEQUENCES: 391
; ; CORRESPONDENCE ADDRESS:
; ; ; ADDRESSEE: Human Genome Sciences, Inc.
; ; ; STREET: 9410 Key West Avenue
; ; ; CITY: Rockville
; ; ; STATE: Maryland
; ; ; COUNTRY: USA
; ; ; ZIP: 20850
; ; COMPUTER READABLE FORM:
; ; ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; ; ; COMPUTER: HP Vectra 486/33
; ; ; OPERATING SYSTEM: MSDOS version 6.2
; ; ; SOFTWARE: ASCII Text
; ; CURRENT APPLICATION DATA:
; ; ; APPLICATION NUMBER: US/08/961,527
; ; ; FILING DATE:
; ; ; CLASSIFICATION: 424
; ; PRIOR APPLICATION DATA:
; ; ; APPLICATION NUMBER:
; ; ; FILING DATE:
; ; ATTORNEY/AGENT INFORMATION:
; ; ; NAME: Brookes, A. Anders
; ; ; REGISTRATION NUMBER: 36,373
; ; ; REFERENCE/DOCKET NUMBER: PB340P1
; ; TELECOMMUNICATION INFORMATION:
; ; ; TELEPHONE: (301) 309-8504
; ; ; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 27:
; ; SEQUENCE CHARACTERISTICS:
; ; ; LENGTH: 6004 base pairs
; ; ; TYPE: nucleic acid
; ; ; STRANDEDNESS: double
; ; ; TOPOLOGY: linear
US-08-961-527-27

Alignment Scores:
Pred. No.:   76.2   Length: 6004
Score: 49.00   Matches: 10
Percent Similarity: 70.59%   Conservative: 2
Best Local Similarity: 58.82%   Mismatches: 5
Query Match: 47.12%   Indels: 0
DB: 4   Gaps: 0

US-09-833-017B-4 (1-21) x US-08-961-527-27 (1-6004)

Qy   5 SerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
       ||||||||| ||| :::|||||||||| ::||||||||||||||
Db 418 TCTACCTTTTCTCCGTTTTTCACTAAAAGTTCACGGCTTCTCTAGGAAAA 468

RESULT 5
US-09-134-001C-302/c
; Sequence 302, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; ; APPLICANT: Lynn Doucette-Stamm et al
; ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; ; FILE REFERENCE: GTC-007
; ; CURRENT APPLICATION NUMBER: US/09/134,001C
; ; CURRENT FILING DATE: 1998-08-13
; ; PRIOR APPLICATION NUMBER: US 60/064,964
; ; PRIOR FILING DATE: 1997-11-08
; ; PRIOR APPLICATION NUMBER: US 60/055,779
; ; PRIOR FILING DATE: 1997-08-14
; ; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 302
; ; LENGTH: 1317
; ; TYPE: DNA
; ; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-302

Alignment Scores:
Pred. No.:   17.9   Length: 1317
Score: 48.00   Matches: 9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 46.15%   Indels: 0
DB: 4   Gaps: 0

US-09-833-017B-4 (1-21) x US-09-134-001C-302 (1-1317)

Qy 7 PhePheArgLeuPheAsnArgSerPhe 15
      ||||||||||||||||||||||||||||
Db 54 TTTTTTCGCTTGTTTAATCGTAGTTTT 28

RESULT 6
US-07-971-819A-1/c
; Sequence 1, Application US/07971819A
; Patent No. 5420029
; GENERAL INFORMATION:
; ; APPLICANT: Gelfand, David H.
; ; APPLICANT: Lawyer, Frances C.
; ; APPLICANT: Stoffel, Susanne
; ; TITLE OF INVENTION: Purified Thermostable Nucleic Acid
; ; TITLE OF INVENTION: Polymerase Enzyme from Thermotoga Maritima
; ; NUMBER OF SEQUENCES: 46
; ; CORRESPONDENCE ADDRESS:
; ; ; ADDRESSEE: Hoffmann-La Roche Inc.
; ; ; STREET: 340 Kingsland Street
; ; ; CITY: Nutley
; ; ; STATE: New Jersey
; ; ; COUNTRY: USA
; ; ; ZIP: 07110-1199
; ; COMPUTER READABLE FORM:
; ; ; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb storage
; ; ; COMPUTER: Apple Macintosh
; ; ; OPERATING SYSTEM: Macintosh 7
; ; ; SOFTWARE: WordPerfect
; ; CURRENT APPLICATION DATA:
; ; ; APPLICATION NUMBER: US/07/971,819A
; ; ; FILING DATE: 19930203
; ; ; CLASSIFICATION: 435
; ; ATTORNEY/AGENT INFORMATION:
; ; ; NAME: Stacey R. Sias, Ph.D.
; ; ; REGISTRATION NUMBER: 32,630
; ; ; REFERENCE/DOCKET NUMBER:
; ; TELECOMMUNICATION INFORMATION:
; ; ; TELEPHONE: (510) 814-2863
; ; ; TELEFAX: (510) 814-2977
; ; ; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; ; SEQUENCE CHARACTERISTICS:
; ; ; LENGTH: 2682 base pairs
; ; ; TYPE: nucleic acid
; ; ; STRANDEDNESS: single
; ; ; TOPOLOGY: linear
; ; MOLECULE TYPE: DNA (genomic)
; ; FEATURE:
; ; ; NAME/KEY: CDS

```
;
    LOCATION:    1..2682
US-07-971-819A-1

Alignment Scores:
Pred. No.:                 78.3                Length:         2682
Score:                     46.50              Matches:        12
Percent Similarity:        57.14%             Conservative:   0
Best Local Similarity:     57.14%             Mismatches:     8
Query Match:               44.71%             Indels:         1
DB:                        1                  Gaps:           1

US-09-833-017B-4 (1-21) x US-07-971-819A-1 (1-2682)

Qy    1 SerGlySerLeuSerThrPhePheArgLeuPhe---AsnArgSerPheThrGlnAlaLeu 19
        ||||||  |||  ||||||  ||||||    |||   |||
Db 1134 TCCGGGGTCCTCCAGAATTTCTTTGAGCTTTTTCAGAACCTCTTTTCGTCCAGGTTCTG 1075

Qy   20 Gly 20
        |||
Db 1074 GGC 1072

RESULT 7
US-07-977-434-3/c
; Sequence 3, Application US/07977434
; Patent No. 5466591
; GENERAL INFORMATION:
    APPLICANT:  Gelfand, David H.
    APPLICANT:  Abramson, Richard D.
    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
    NUMBER OF SEQUENCES:  38
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Hoffmann-La Roche Inc.
      STREET:  340 Kingsland Street
      CITY:  Nutley
      STATE:  New Jersey
      ZIP:  07110-1199
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  Macintosh
      OPERATING SYSTEM:  7
      SOFTWARE:  WordPerfect 2.1
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/07/977,434
      FILING DATE:
      CLASSIFICATION:  435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 590,490
      FILING DATE:  28-SEP-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 590,466
      FILING DATE:  28-SEP-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 590,213
      FILING DATE:  28-SEP-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 523,394
      FILING DATE:  15-MAY-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 143,441
      FILING DATE:  12-JAN-1988
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 063,509
      FILING DATE:  17-JUN-1987
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 899,241
      FILING DATE:  22-AUG-1986
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 746,121
      FILING DATE:  15-AUG-1991
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  WO PCT/US90/07641
      FILING DATE:  21-DEC-1990

    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 585,471
      FILING DATE:  20-SEP-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 455,611
      FILING DATE:  22-DEC-1989
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 609,157
      FILING DATE:  02-NOV-1990
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US 557,517
      FILING DATE:  24-JUL-1990
    ATTORNEY/AGENT INFORMATION:
      NAME:  Luann Cserr
      REGISTRATION NUMBER:  31,822
      REFERENCE/DOCKET NUMBER:  Case No. 5466591 8753
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (510) 814-2972
INFORMATION FOR SEQ ID NO:  3:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  2682 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
      MOLECULE TYPE:  DNA (genomic)
      HYPOTHETICAL:  NO
      ANTI-SENSE:  NO
    ORIGINAL SOURCE:
      ORGANISM:  Thermotoga maritima
    FEATURE:
      NAME/KEY:  CDS
      LOCATION:  1..2679
;
US-07-977-434-3

Alignment Scores:
Pred. No.:                 78.3                Length:         2682
Score:                     46.50              Matches:        12
Percent Similarity:        57.14%             Conservative:   0
Best Local Similarity:     57.14%             Mismatches:     8
Query Match:               44.71%             Indels:         1
DB:                        1                  Gaps:           1

US-09-833-017B-4 (1-21) x US-07-977-434-3 (1-2682)

Qy    1 SerGlySerLeuSerThrPhePheArgLeuPhe---AsnArgSerPheThrGlnAlaLeu 19
        ||||||  |||  ||||||  ||||||    |||   |||
Db 1134 TCCGGGGTCCTCCAGAATTTCTTTGAGCTTTTTCAGAACCTCTTTTCGTCCAGGTTCTG 1075

Qy   20 Gly 20
        |||
Db 1074 GGC 1072

RESULT 8
US-08-475-231-1/c
; Sequence 1, Application US/08475231
; Patent No. 5624833
; GENERAL INFORMATION:
    APPLICANT:  Gelfand, David H.
    APPLICANT:  Lawyer, Frances C.
    APPLICANT:  Stoffel, Susanne
    TITLE OF INVENTION:  Purified Thermostable Nucleic Acid
    TITLE OF INVENTION:  Polymerase Enzyme from Thermotoga Maritima
    NUMBER OF SEQUENCES:  46
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Hoffmann-La Roche Inc.
      STREET:  340 Kingsland Street
      CITY:  Nutley
      STATE:  New Jersey
      COUNTRY:  USA
      ZIP:  07110-1199
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette, 3.50 inch, 800 Kb storage
      COMPUTER:  Apple Macintosh
;
```

```
; OPERATING SYSTEM: Macintosh 7
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,231
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,819
; FILING DATE: 03-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stacey R. Sias, Ph.D.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2682 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2682
US-08-475-231-1

Alignment Scores:
Pred. No.:                78.3        Length:          2682
Score:                    46.50       Matches:         12
Percent Similarity:       57.14%      Conservative:    0
Best Local Similarity:    57.14%      Mismatches:      8
Query Match:              44.71%      Indels:          1
DB:                       1           Gaps:            1

US-09-833-017B-4 (1-21) x US-08-475-231-1 (1-2682)

Qy     1 SerGlySerLeuSerThrPheArgLeuPhe--AsnArgSerPheThrGlnAlaLeu 19
         ||||||  ||| |||||| |||||| |||  |||||| |||  |||
Db  1134 TCCGGGTCCTCCAGAATTCTTTGAGCTTTTTCAGAACCTCTTTTCGTCCAGGTTCTG 1075

Qy    20 Gly 20
         |||
Db  1074 GGC 1072

RESULT 9
US-08-458-819-3/c
; Sequence 3, Application US/08458819
; Patent No. 5795762
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,819
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/977,434
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2682 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2679
US-08-458-819-3

Alignment Scores:
Pred. No.:                78.3        Length:          2682
Score:                    46.50       Matches:         12
Percent Similarity:       57.14%      Conservative:    0
Best Local Similarity:    57.14%      Mismatches:      8
Query Match:              44.71%      Indels:          1
DB:                       1           Gaps:            1

US-09-833-017B-4 (1-21) x US-08-458-819-3 (1-2682)

Qy     1 SerGlySerLeuSerThrPheArgLeuPhe--AsnArgSerPheThrGlnAlaLeu 19
```

Db   1134 TCCGGGGTCCTCCAGAATTCTTTGAGCTTTTTTCAGAACCTCTTTTTCGTCCAGGTTCTG 1075

Qy   20 Gly 20

Db   1074 GGC 1072

RESULT 10
US-09-105-697-9/c
; Sequence 9, Application US/09105697
; Patent No. 6228628
; GENERAL INFORMATION:
;   APPLICANT: Gelfand Ph.D., David H.
;   APPLICANT: Reichert, Fred L.
;   TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Roche Molecular Systems
;   STREET: 1080 U.S. Highway 202
;   CITY: Branchburg
;   STATE: New Jersey
;   COUNTRY: United States
;   ZIP: 08876
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/105,697
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Petry Ph.D., Douglas A.
;   REGISTRATION NUMBER: 35321
;   REFERENCE/DOCKET NUMBER: 1043
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510)814-2974
;   TELEFAX: (510)814-2977
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2682 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-09-105-697-9

Alignment Scores:
Pred. No.:   Length: 2682
Score: 78.3   Matches: 12
Percent Similarity: 46.50   Conservative: 0
Best Local Similarity: 57.14%   Mismatches: 8
Query Match: 44.71%   Indels: 1
DB: 4   Gaps: 1

US-09-833-017B-4 (1-21) x US-09-105-697-9 (1-2682)

Qy   1 SerGlySerLeuSerThrPhePheArgLeuPhe--AsnArgSerPheThrGlnAlaLeu 19
Db 1134 TCCGGGGGTCCTCCAGAATTCTTTGAGCTTTTTTCAGAACCCTCTTTTTCGTCCAGGTTCTG 1075
Qy   20 Gly 20
Db 1074 GGC 1072

RESULT 11
PCT-US91-07035-3/c
; Sequence 3, Application PC/TUS9107035
; GENERAL INFORMATION:
;   APPLICANT: Gelfand, David H.
;   APPLICANT: Abramson, Richard D.
;   TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF

;   TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
;   NUMBER OF SEQUENCES: 38
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Cetus Corporation
;   STREET: 1400 Fifty-third Street
;   CITY: Emeryville
;   STATE: California
;   ZIP: 94608
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WordPerfect 5.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US91/07035
;   FILING DATE: 19910930
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 590,490
;   FILING DATE: 28-SEP-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 590,466
;   FILING DATE: 28-SEP-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 590,213
;   FILING DATE: 28-SEP-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 523,394
;   FILING DATE: 15-MAY-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 143,441
;   FILING DATE: 12-JAN-1988
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 063,509
;   FILING DATE: 17-JUN-1987
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 899,241
;   FILING DATE: 22-AUG-1986
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 746,121
;   FILING DATE: 15-AUG-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/US90/07641
;   FILING DATE: 21-DEC-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 585,471
;   FILING DATE: 20-SEP-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 455,611
;   FILING DATE: 22-DEC-1989
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 609,157
;   FILING DATE: 02-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 557,517
;   FILING DATE: 24-JUL-1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sias Ph.D, Stacey R.
;   REGISTRATION NUMBER: 32,630
;   REFERENCE/DOCKET NUMBER: Case No. 2580
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-420-3300
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2682 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Thermotoga maritima

```
;   FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 1..2679
PCT-US91-07035-3

Alignment Scores:
Pred. No.:                  78.3                    Length:        2682
Score:                      46.50                   Matches:       12
Percent Similarity:         57.14%                  Conservative:  0
Best Local Similarity:      57.14%                  Mismatches:    8
Query Match:                44.71%                  Indels:        1
DB:                                                 Gaps:          1

US-09-833-017B-4 (1-21) x PCT-US91-07035-3 (1-2682)

Qy     1 SerGlySerLeuSerThrPheArgLeuPhe---AsnArgSerPheThrGlnAlaLeu 19
           |||||||   |||    ||||||   |||    ||||   |||   |||
Db  1134 TCCGGGTCCTCCAGAATTTCTTTGAGCTTTTTCAGAACCTCTTTTCGTCCAGGTTCTG 1075

Qy    20 Gly 20
           |||
Db  1074 GGC 1072

RESULT 12
US-09-484-970B-17/c
; Sequence 17, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
;   APPLICANT: Jones, Karen A.
;   APPLICANT: Volkmuth, Wayne
;   APPLICANT: Walker, Michael G.
;   TITLE OF INVENTION: BONE REMODELING GENES
;   FILE REFERENCE: PB-0014 US
;   CURRENT APPLICATION NUMBER: US/09/484,970B
;   CURRENT FILING DATE: 2000-01-18
;   NUMBER OF SEQ ID NOS: 172
;   SOFTWARE: PERL Program
;   SEQ ID NO 17
;       LENGTH: 1651
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;   FEATURE:
;       NAME/KEY: misc_feature
;       OTHER INFORMATION: Incyte ID No. 6426186 126510.2CB1
;       NAME/KEY: unsure
;       LOCATION: 767-846
;       OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-17

Alignment Scores:
Pred. No.:                  79.4                    Length:        1651
Score:                      45.00                   Matches:       9
Percent Similarity:         76.92%                  Conservative:  1
Best Local Similarity:      69.23%                  Mismatches:    3
Query Match:                43.27%                  Indels:        0
DB:                                                 Gaps:          0

US-09-833-017B-4 (1-21) x US-09-484-970B-17 (1-1651)

Qy     4 LeuSerThrPheArgLeuPheAsnArgSerPheThr 16
           :::|||||||||||||    |||   ||||||   |||
Db   519 GTGAGCACCCTTCTCCGGTTCTTCCCCGCTCCTTGACC 481

RESULT 13
US-08-961-527-16/c
; Sequence 16, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;   NUMBER OF SEQUENCES: 391
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Human Genome Sciences, Inc.
```

```
;       STREET: 9410 Key West Avenue
;       CITY: Rockville
;       STATE: Maryland
;       COUNTRY: USA
;       ZIP: 20850
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER: HP Vectra 486/33
;       OPERATING SYSTEM: MSDOS version 6.2
;       SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/961,527
;       FILING DATE:
;       CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Brookes, A. Anders
;       REGISTRATION NUMBER: 36,373
;       REFERENCE/DOCKET NUMBER: PB340P1
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (301) 309-8504
;       TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 8411 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
US-08-961-527-16

Alignment Scores:
Pred. No.:                  581                     Length:        8411
Score:                      45.00                   Matches:       8
Percent Similarity:         81.25%                  Conservative:  5
Best Local Similarity:      50.00%                  Mismatches:    3
Query Match:                43.27%                  Indels:        0
DB:                                                 Gaps:          0

US-09-833-017B-4 (1-21) x US-08-961-527-16 (1-8411)

Qy     6 ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
           ||||||:|||:||||||    :::::|||||||:::::||||||:::||||::::||
Db  3443 ACCTTCTATCGTTTAGATACCAAAACTTTCACCGAGGCAATTCAAAAA 3396

RESULT 14
US-09-784-316-3/c
; Sequence 3, Application US/09784316
; Patent No. 6461843
; GENERAL INFORMATION:
;   APPLICANT: WEI, Ming-Hui et al.
;   TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001139
;   CURRENT APPLICATION NUMBER: US/09/784,316
;   CURRENT FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 5
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 3
;       LENGTH: 65042
;       TYPE: DNA
;       ORGANISM: Human
;   FEATURE:
;       NAME/KEY: misc_feature
;       LOCATION: (1)...(65042)
;       OTHER INFORMATION: n = A,T,C or G
US-09-784-316-3

Alignment Scores:
Pred. No.:                  7.07e+03                Length:        65042
Score:                      45.00                   Matches:       8
```

```
Percent Similarity:      72.22%        Conservative:   5
Best Local Similarity:   44.44%        Mismatches:     5
Query Match:             43.27%        Indels:         0
DB:                      4             Gaps:           0

US-09-833-017B-4 (1-21) x US-09-784-316-3 (1-65042)

Qy      1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAla 18
          ||||||  |||||      :::|||::: |||   :::::::|||||||:::
Db  42098 TCAGGAATTCTATCAAATGTTTACAGAATAATTATACCAATTATACACAATCT 42045

RESULT 15
US-08-745-995A-22/c
; Sequence 22, Application US/08745995A
; Patent No. 6372887
  GENERAL INFORMATION:
     APPLICANT:  Silver, Gary M.
     APPLICANT:  Wisnewski, Nancy
     TITLE OF INVENTION:  No. 6372887el Serine Protease Inhibitor
     TITLE OF INVENTION:  Nucleic Acid Molecules, Proteins
     TITLE OF INVENTION:  and Uses Thereof
     NUMBER OF SEQUENCES:  41
     CORRESPONDENCE ADDRESS:
        ADDRESSEE:  Heska Corporation
        STREET:  1825 Sharp Point Drive
        CITY:  Fort Collins
        STATE:  Colorado
        COUNTRY:  USA
        ZIP:  80525
     COMPUTER READABLE FORM:
        MEDIUM TYPE:  Floppy disk
        COMPUTER:  IBM PC compatible
        OPERATING SYSTEM:  Windows 95
        SOFTWARE:  WordPerfect for Windows, Version 7.0
     CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  US/08/745,995A
        FILING DATE:  07-NOV-1996
        CLASSIFICATION:  536
     ATTORNEY/AGENT INFORMATION:
        NAME:  Verser, Carol Talkington
        REGISTRATION NUMBER:  37,459
        REFERENCE/DOCKET NUMBER:  FC-2
     TELECOMMUNICATION INFORMATION:
        TELEPHONE:  970/493-7272
        TELEFAX:  970/484-9505
  INFORMATION FOR SEQ ID NO:  22:
     SEQUENCE CHARACTERISTICS:
        LENGTH:  1179 nucleotides
        TYPE:  nucleic acid
        STRANDEDNESS:  single
        TOPOLOGY:  linear
     MOLECULE TYPE:  cDNA
US-08-745-995A-22

Alignment Scores:
Pred. No.:                              Length:          1179
Score:                   78.9           Matches:         10
Percent Similarity:      44.00          Conservative:    0
Best Local Similarity:   66.67%         Mismatches:      5
Query Match:             66.67%         Indels:          0
DB:                      42.31%         Gaps:            0
                         4

US-09-833-017B-4 (1-21) x US-08-745-995A-22 (1-1179)

Qy      1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPhe 15
          ||| |||  |||||||||||||    ||||||     |||||
Db   1041 AGCTGCAGCTTCAGCACCTTCTTCATTTACTTCAATGAAAGCTTT 997
```

Search completed: November 8, 2002, 23:41:22
Job time :28.9403 secs

```
                GenCore version 5.1.3
        Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      November  8, 2002, 22:57:42 ; Search time 20.6866 Seconds
                                           (without alignments)
                                           360.161 Million cell updates/sec

Title:        US-09-833-017B-4
Perfect score: 104
Sequence:     1 SGSLSTFFRLFNRSFTQALGK 21

Scoring table: BLOSUM62
              Xgapop  10.0 , Xgapext   0.5
              Ygapop  10.0 , Ygapext   0.5
              Fgapop   6.0 , Fgapext   7.0
              Delop    6.0 , Delext    7.0

Searched:     320260 seqs, 177332727 residues

Total number of hits satisfying chosen parameters:        640520

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US09833017/runat_05112002_105351_4940/app_query.fasta_1.398
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09833017_@CGN_1_1_18_@runat_05112002_105351_4940
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    Published_Applications_NA:*
        1: /ptodata/1/pubpna/US07_PUBCOMB.seq:*
        2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
        3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
        4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
        5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
        6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
        7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
        8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
        9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
       10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
       11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
       12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
       13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
       14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES

Result    Query
  No.  Score Match Length DB   ID              Description
------ ----- ----- ------ -- ----------------  ---------------------------
     1   104 100.0    141 10 US-09-833-017-1   Sequence 1, Appli
     2   104 100.0   2557 10 US-09-833-017-21  Sequence 21, Appl
   c 3    52  50.0  10476 10 US-09-964-824A-98 Sequence 98, Appl
   c 4    52  50.0  10476 10 US-09-964-824A-552 Sequence 552, App
```

```
    5   50 48.1  2017 12 US-10-007-693-12    Sequence 12, Appl
  c 6   49 47.1   694 10 US-09-765-272-175   Sequence 175, App
  c 7   47 45.2   870 10 US-09-770-445-564   Sequence 564, App
  c 8   46 44.2 48667 10 US-09-822-268A-3    Sequence 3, Appli
  c 9   45 43.3   284 10 US-09-864-761-32659 Sequence 32659, A
   10   45 43.3   333 10 US-09-917-800A-276  Sequence 276, App
   11   45 43.3   430 10 US-09-728-446-245   Sequence 245, App
  c 12  45 43.3   489 10 US-09-733-387-11    Sequence 11, Appl
   13   45 43.3   525 10 US-09-974-300-6233  Sequence 6233, Ap
   14   45 43.3   528 10 US-09-733-387-31    Sequence 31, Appl
   15   45 43.3   584 10 US-09-864-761-16165 Sequence 16165, A
   16   45 43.3   612 10 US-09-733-387-51    Sequence 51, Appl
   17   45 43.3   702 10 US-09-733-387-9     Sequence 9, Appli
   18   45 43.3   741 10 US-09-733-387-29    Sequence 29, Appl
   19   45 43.3   825 10 US-09-733-387-49    Sequence 49, Appl
   20   45 43.3   897 10 US-09-733-387-5     Sequence 5, Appli
   21   45 43.3   936 10 US-09-733-387-25    Sequence 25, Appl
   22   45 43.3   942 10 US-09-974-300-2095  Sequence 2095, Ap
   23   45 43.3  1020 10 US-09-733-387-45    Sequence 45, Appl
   24   45 43.3  1080 10 US-09-733-387-7     Sequence 7, Appli
   25   45 43.3  1119 10 US-09-733-387-27    Sequence 27, Appl
   26   45 43.3  1203 10 US-09-733-387-47    Sequence 47, Appl
   27   45 43.3  1527 10 US-09-733-387-3     Sequence 3, Appli
  c 28  45 43.3  1539 10 US-09-887-377-1     Sequence 1, Appli
   29   45 43.3  1566 10 US-09-733-387-23    Sequence 23, Appl
   30   45 43.3  1650 10 US-09-733-387-43    Sequence 43, Appl
   31   45 43.3  4036 10 US-09-733-387-53    Sequence 53, Appl
   32   45 43.3  5959 10 US-09-954-456-1996  Sequence 1996, Ap
   33   44 42.3   348 10 US-09-770-791-742   Sequence 742, App
  c 34  44 42.3   599 10 US-09-864-761-12681 Sequence 12681, A
   35   44 42.3  1053 10 US-09-833-381-365   Sequence 365, App
   36   44 42.3  1405 10 US-09-943-272-1     Sequence 1, Appli
  c 37  44 42.3 32248 10 US-09-764-864-1769  Sequence 1769, Ap
   38   44 42.3 32248 10 US-09-764-877-3487  Sequence 3487, Ap
  c 39  44 42.3 397658 10 US-09-813-320-3    Sequence 3, Appli
   40   44 42.3 640681 10 US-09-790-988-1    Sequence 1, Appli
  c 41  43.5 41.8   438 10 US-09-960-352-9083 Sequence 9083, Ap
  c 42  43 41.3   215 10 US-09-815-242-2813  Sequence 2813, Ap
   43   43 41.3   344 10 US-09-815-242-2237  Sequence 2237, Ap
  c 44  43 41.3   439 10 US-09-563-817-724   Sequence 724, App
   45   43 41.3   501 10 US-09-815-242-4689  Sequence 4689, Ap


                            ALIGNMENTS

RESULT 1
US-09-833-017-1
; Sequence 1, Application US/09833017
; Patent No. US2002008130201 
; Patent No. US20020081302A1
; GENERAL INFORMATION:
;  APPLICANT: CVITKOVITCH, Dennis
;  TITLE OF INVENTION: SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATME
;  TITLE OF INVENTION: OF CARIES
;  FILE REFERENCE: P0485US1
;  CURRENT APPLICATION NUMBER: US/09/833,017
;  CURRENT FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: PCT/CA00/00605
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 2,302,861
;  PRIOR FILING DATE: 2000-04-10
;  PRIOR APPLICATION NUMBER: 2,332,733
;  PRIOR FILING DATE: 2001-02-20
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: Patentin version 3.0
;  SEQ ID NO 1
;  LENGTH: 141
;  TYPE: DNA
;  ORGANISM: Streptococcus mutans
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(141)
US-09-833-017-1
```

```
Alignment Scores:
Pred. No.:                     1.62e-10      Length:            141
Score:                         104.00        Matches:           21
Percent Similarity:            100.00%       Conservative:      0
Best Local Similarity:         100.00%       Mismatches:        0
Query Match:                   100.00%       Indels:            0
DB:                            10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-833-017-1 (1-141)

Qy      1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76  AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 135

Qy     21  Lys 21
           |||
Db    136  AAA 138

RESULT 2
US-09-833-017-21
; Sequence 21, Application US/09833017
; Patent No. US2002081302A1
; GENERAL INFORMATION:
;   APPLICANT: CVITKOVITCH, Dennis
;   TITLE OF INVENTION: SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT
;   TITLE OF INVENTION: OF CARIES
;   FILE REFERENCE: P04885US1
;   CURRENT APPLICATION NUMBER: US/09/833,017
;   CURRENT FILING DATE: 2001-04-10
;   PRIOR APPLICATION NUMBER: PCT/CA00/00605
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: 2,302,861
;   PRIOR FILING DATE: 2000-04-10
;   PRIOR APPLICATION NUMBER: 2,332,733
;   PRIOR FILING DATE: 2001-02-20
;   NUMBER OF SEQ ID NOS: 28
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
;   LENGTH: 2557
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2557)
US-09-833-017-21

Alignment Scores:
Pred. No.:                     5.55e-09      Length:            2557
Score:                         104.00        Matches:           21
Percent Similarity:            100.00%       Conservative:      0
Best Local Similarity:         100.00%       Mismatches:        0
Query Match:                   100.00%       Indels:            0
DB:                            10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-833-017-21 (1-2557)

Qy      1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    176  AGCGGAAGCCTATCAACATTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGGA 235

Qy     21  Lys 21
           |||
Db    236  AAA 238

RESULT 3
US-09-964-824A-98/c
; Sequence 98, Application US/09964824A
; Patent No. US2002102531A1
; GENERAL INFORMATION:
;   APPLICANT: Horrigan, Stephen
;   TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
;   TITLE OF INVENTION: Sets
;   FILE REFERENCE: 689290-73
```

```
;   CURRENT APPLICATION NUMBER: US/09/964,824A
;   CURRENT FILING DATE: 2001-09-27
;   PRIOR APPLICATION NUMBER: US/60/236,033
;   PRIOR FILING DATE: 2000-09-28
;   PRIOR APPLICATION NUMBER: US/60/236,032
;   PRIOR FILING DATE: 2000-09-28
;   PRIOR APPLICATION NUMBER: US/60/236,028
;   PRIOR FILING DATE: 2000-09-28
;   NUMBER OF SEQ ID NOS: 583
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 98
;   LENGTH: 10476
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-964-824A-98

Alignment Scores:
Pred. No.:                     39.5          Length:            10476
Score:                         52.00         Matches:           10
Percent Similarity:            76.47%        Conservative:      3
Best Local Similarity:         58.82%        Mismatches:        4
Query Match:                   50.00%        Indels:            0
DB:                            10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-964-824A-98 (1-10476)

Qy      1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
           |||||||||:   |||:||||   |||        ||  :||::||||||||
Db    543  TCTGGTTCTATGGACACATGGTTTGCATTAGTCAGAAGAGCATTTACACAA 493

RESULT 4
US-09-964-824A-552/c
; Sequence 552, Application US/09964824A
; Patent No. US2002102531A1
; GENERAL INFORMATION:
;   APPLICANT: Horrigan, Stephen
;   TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;   TITLE OF INVENTION: Sets
;   FILE REFERENCE: 689290-73
;   CURRENT APPLICATION NUMBER: US/09/964,824A
;   CURRENT FILING DATE: 2001-09-27
;   PRIOR APPLICATION NUMBER: US/60/236,033
;   PRIOR FILING DATE: 2000-09-28
;   PRIOR APPLICATION NUMBER: US/60/236,032
;   PRIOR FILING DATE: 2000-09-28
;   PRIOR APPLICATION NUMBER: US/60/236,028
;   PRIOR FILING DATE: 2000-09-28
;   NUMBER OF SEQ ID NOS: 583
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 552
;   LENGTH: 10476
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-964-824A-552

Alignment Scores:
Pred. No.:                     39.5          Length:            10476
Score:                         52.00         Matches:           10
Percent Similarity:            76.47%        Conservative:      3
Best Local Similarity:         58.82%        Mismatches:        4
Query Match:                   50.00%        Indels:            0
DB:                            10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-964-824A-552 (1-10476)

Qy      1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
           |||||||||:   |||:||||   |||        ||  :||::||||||||
Db    543  TCTGGTTCTATGGACACATGGTTTGCATTAGTCAGAAGAGCATTTACACAA 493

RESULT 5
US-10-007-693-12
; Sequence 12, Application US/10007693
; Patent No. US2002146776A1
```

; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 12
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis
US-10-007-693-12

Alignment Scores:

Pred. No.:	11.9	Length:	2017
Score:	50.00	Matches:	10
Percent Similarity:	78.57%	Conservative:	1
Best Local Similarity:	71.43%	Mismatches:	3
Query Match:	48.08%	Indels:	0
DB:	12	Gaps:	0

US-09-833-017B-4 (1-21) x US-10-007-693-12 (1-2017)

```
Qy    3 SerLeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
          ::::|||  ||||||||| ||||||||||||||  |||
Db  353 GCACCAATGACCTTTTCAGTTTATTTAATAGGTCTTTAACA 394
```

RESULT 6
US-09-765-272-175/c
; Sequence 175, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB34QP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 175:
US-09-765-272-175

Alignment Scores:

Pred. No.:	4.82	Length:	694
Score:	49.00	Matches:	10
Percent Similarity:	70.59%	Conservative:	2
Best Local Similarity:	58.82%	Mismatches:	5
Query Match:	47.12%	Indels:	0
DB:	10	Gaps:	0

US-09-833-017B-4 (1-21) x US-09-765-272-175 (1-694)

```
Qy    5 SerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
          |||||||||       |||     ::||||||||   ::|||||||||
Db  210 TCTACCTTTTCTCCGTTTTTCACTAAAAGTTTCACGGCTTCTCTAGGAAAA 160
```

RESULT 7
US-09-770-445-564
; Sequence 564, Application US/09770445
; Patent No. US2002023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SEQ ID NO 564
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-564

Alignment Scores:

Pred. No.:	14.2	Length:	870
Score:	47.00	Matches:	9
Percent Similarity:	65.00%	Conservative:	4
Best Local Similarity:	45.00%	Mismatches:	7
Query Match:	45.19%	Indels:	0
DB:	10	Gaps:	0

US-09-833-017B-4 (1-21) x US-09-770-445-564 (1-870)

```
Qy    2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
          |||  :::  |||   |||  ||||||  :|||||||:::  ||||||:::  |||||
Db  785 GGTTTATCCAAACTCTTTTGTTCTTTTTCGCCAGACATTTCAGAACCTTTCGGAAAA 844
```

RESULT 8
US-09-822-268A-3/c
; Sequence 3, Application US/09822268A
; Patent No. US20020048787A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL000667
CURRENT APPLICATION NUMBER: US/09/822,268A
CURRENT FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/211,387
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
PRIOR FILING DATE: 2001-02-13
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
  LENGTH: 48667
  TYPE: DNA
  ORGANISM: HUMAN
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)...(48667)
  OTHER INFORMATION: n = A,T,C or G
US-09-822-268A-3

Alignment Scores:
Pred. No.:   2.89e+03   Length: 48667
Score: 46.00   Matches: 9
Percent Similarity: 72.22%   Conservative: 4
Best Local Similarity: 50.00%   Mismatches: 5
Query Match: 44.23%   Indels: 0
DB: 10   Gaps: 0

US-09-833-017B-4 (1-21) x US-09-822-268A-3 (1-48667)

Qy   2 GlySerLeuThrPhePheAsnArgSerPheThrGlnAlaLeu 19
           : :   :: || ::: |||::: |||||:::
Db 15206 GGGAGTCTCTCACGTTGCTCAAGCTGATCTCAAATTCCTGGGCTCAAGCGATC 15153

RESULT 9
US-09-864-761-32659/c
; Sequence 32659, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
  APPLICANT: Rank, David R.
  APPLICANT: Hanzel, David K.
  APPLICANT: Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aeomica-X-1
  CURRENT APPLICATION NUMBER: US/09/864,761
  CURRENT FILING DATE: 2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: PCT/US01/00666
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00667
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00664
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00669
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00665
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00668
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00663
  PRIOR FILING DATE: 2001-01-30

  PRIOR APPLICATION NUMBER: PCT/US01/00662
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00661
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00670
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: US 60/234,687
  PRIOR FILING DATE: 2000-09-21
  PRIOR APPLICATION NUMBER: US 09/608,408
  PRIOR FILING DATE: 2000-06-30
  PRIOR APPLICATION NUMBER: US 09/774,203
  PRIOR FILING DATE: 2001-01-29
  NUMBER OF SEQ ID NOS: 49117
  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32659
  LENGTH: 284
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  OTHER INFORMATION: MAP TO AC018552.2
  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
  OTHER INFORMATION: NT HIT: D21877.1, EVALUE 1.70e-01
  OTHER INFORMATION: EST_HUMAN HIT: AW592568.1, EVALUE 8.00e-54
  OTHER INFORMATION: SWISSPROT HIT: Q91062, EVALUE 1.70e+00
US-09-864-761-32659

Alignment Scores:
Pred. No.:   8.13   Length: 284
Score: 45.00   Matches: 9
Percent Similarity: 76.92%   Conservative: 1
Best Local Similarity: 69.23%   Mismatches: 3
Query Match: 43.27%   Indels: 0
DB: 10   Gaps: 0

US-09-833-017B-4 (1-21) x US-09-864-761-32659 (1-284)

Qy   4 LeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
           :::|||||||||||||||||||| || :::|||::: |||
Db 177 GTGAGCACCTTCTTCCGGTTCTTCCCCGTCCTTGACC 139

RESULT 10
US-09-917-800A-276/c
; Sequence 276, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
  APPLICANT: Mendrick, Donna
  APPLICANT: Porter, Mark
  APPLICANT: Johnson, Kory
  APPLICANT: Castle, Arthur
  APPLICANT: Elashoff, Michael
  APPLICANT: Gene Logic, Inc.
  TITLE OF INVENTION: Molecular Toxicology Modeling
  FILE REFERENCE: 44921-5038-US
  CURRENT APPLICATION NUMBER: US/09/917,800A
  CURRENT FILING DATE: 2001-07-31
  PRIOR APPLICATION NUMBER: US 60/222,040
  PRIOR FILING DATE: 2000-07-31
  PRIOR APPLICATION NUMBER: US 60/222,880
  PRIOR FILING DATE: 2000-11-02
  PRIOR APPLICATION NUMBER: US 60/290,029
  PRIOR FILING DATE: 2001-05-11
  PRIOR APPLICATION NUMBER: US 60/290,645
  PRIOR FILING DATE: 2001-05-15
  PRIOR APPLICATION NUMBER: US 60/292,336
  PRIOR FILING DATE: 2001-05-22
  PRIOR APPLICATION NUMBER: US 60/295,798
  PRIOR FILING DATE: 2001-06-06
  PRIOR APPLICATION NUMBER: US 60/297,457
  PRIOR FILING DATE: 2001-06-13
  PRIOR APPLICATION NUMBER: US 60/298,884
  PRIOR FILING DATE: 2001-06-19
  PRIOR APPLICATION NUMBER: US 60/303,459
  PRIOR FILING DATE: 2001-07-09

```
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 276
;     LENGTH: 333
;     TYPE: DNA
;     ORGANISM: Rattus norvegicus
;     FEATURE:
;     OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA899635
US-09-917-800A-276

Alignment Scores:
Pred. No.:              9.87          Length:            333
Score:                  45.00         Matches:           9
Percent Similarity:     90.91%        Conservative:      1
Best Local Similarity:  81.82%        Mismatches:        1
Query Match:            43.27%        Indels:            0
DB:                     10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-917-800A-276 (1-333)

Qy    1 SerGlySerLeuSerThrPhePheArgLeuPhe 11
          :::||||||||||||||||| ||||||||
Db  210 ACGGGATCTTTATCCACGTTTCTTCGTCTTTC 178

RESULT 11
US-09-728-446-245
; Sequence 245, Application US/09728446
; Patent No. US2002008168A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: US20020081668Alel Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0101-USA
; CURRENT APPLICATION NUMBER: US/09/728,446
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,270
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1461
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 245
;     LENGTH: 430
;     TYPE: DNA
;     ORGANISM: Mus musculus
US-09-728-446-245

Alignment Scores:
Pred. No.:              13.5          Length:            430
Score:                  45.00         Matches:           11
Percent Similarity:     66.67%        Conservative:      3
Best Local Similarity:  52.38%        Mismatches:        5
Query Match:            43.27%        Indels:            2
DB:                     10            Gaps:              1

US-09-833-017B-4 (1-21) x US-09-728-446-245 (1-430)

Qy    2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThr-----GlnAlaAlaLeu 19
          ||| ||| |||||||||| :::::||     |||    ||| |||
Db  100 GGCGACCTGGACACTTTTTTCCCCTGTTAGTCGTCCCTGGTCCCAGCTGACAGAACCTT 159

Qy   20 Gly 20
          |||
Db  160 GGC 162

RESULT 12
US-09-733-387-11/c
; Sequence 11, Application US/09733387
; Patent No. US20020103359A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
```

```
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020103359Alel Human Membrane Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0104-USA
; CURRENT APPLICATION NUMBER: US/09/733,387
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,427
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;     LENGTH: 489
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-733-387-11

Alignment Scores:
Pred. No.:              15.8          Length:            489
Score:                  45.00         Matches:           9
Percent Similarity:     76.92%        Conservative:      1
Best Local Similarity:  69.23%        Mismatches:        3
Query Match:            43.27%        Indels:            0
DB:                     10            Gaps:              0

US-09-733-387-11 (1-21) x US-09-733-387-11 (1-489)

Qy    4 LeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
          :::||||||||||||||||| ||| ||||| |||
Db  395 GTGAGCACCTTCTTCCGGTTCTTCCCCGCTCCTTGACC 357

RESULT 13
US-09-974-300-6233/c
; Sequence 6233, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6233
;     LENGTH: 525
;     TYPE: DNA
;     ORGANISM: Bacillus clausii
;     FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)...(525)
;     OTHER INFORMATION: n = A,T,C or G
US-09-974-300-6233

Alignment Scores:
Pred. No.:              17.2          Length:            525
Score:                  45.00         Matches:           9
Percent Similarity:     65.00%        Conservative:      4
Best Local Similarity:  45.00%        Mismatches:        7
Query Match:            43.27%        Indels:            0
DB:                     10            Gaps:              0

US-09-833-017B-4 (1-21) x US-09-974-300-6233 (1-525)

Qy    2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
          |||||||| ::::   |||  :::: ||   ||||||||| ::::::||||||
Db  393 GGTTCTCTGGACGGTTTCTTCAAATGAGTAATCTTCCGTATACGCGGGGCCCTGAATAAG 334
```

RESULT 14
US-09-733-387-31/c
; Sequence 31, Application US/09733387
; Patent No. US2002103359A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020103359A1el Human Membrane Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0104-USA
; CURRENT APPLICATION NUMBER: US/09/733,387
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,427
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-733-387-31

Alignment Scores:
Pred. No.:   17.3   Length: 528
Score: 45.00   Matches: 9
Percent Similarity: 76.92%   Conservative: 1
Best Local Similarity: 69.23%   Mismatches: 3
Query Match: 43.27%   Indels: 0
DB: 10   Gaps: 0

US-09-833-017B-4 (1-21) x US-09-733-387-31 (1-528)

Qy   4 LeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
            :::|||||||||||||| ||| ||| ||
Db   395 GTGAGCACCTTCTTCCGGTTCTTCCCCGCTCCTTGACC 357

RESULT 15
US-09-864-761-16165/c
; Sequence 16165, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16165
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018552.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
US-09-864-761-16165

Alignment Scores:
Pred. No.:   19.6   Length: 584
Score: 45.00   Matches: 9
Percent Similarity: 76.92%   Conservative: 1
Best Local Similarity: 69.23%   Mismatches: 3
Query Match: 43.27%   Indels: 0
DB: 10   Gaps: 0

US-09-833-017B-4 (1-21) x US-09-864-761-16165 (1-584)

Qy   4 LeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
            :::|||||||||||||| ||| ||| ||
Db   435 GTGAGCACCTTCTTCCGGTTCTTCCCCGCTCCTTGACC 397

Search completed: November 9, 2002, 00:58:50
Job time : 27.6866 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: November 8, 2002, 22:01:57 ; Search time 825.582 Seconds
                                     (without alignments)
                                     411.958 Million cell updates/sec

Title: US-09-833-017B-4
Perfect score: 104
Sequence: 1 SGSLSTFFRLFNRSTFQALGK 21

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop   6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0

Searched: 1615066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09833017/runat_05112002_105349_4846/app-query.fasta_1.398
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09833017 @CGN_1_1_899 @runat_05112002_105349_4846 -NCPU=6 -ICPU=3.
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
       1:   em_estba:.*
       2:   em_esthum:.*
       3:   em_estin:.*
       4:   em_estmu:.*
       5:   em_estov:.*
       6:   em_estpl:.*
       7:   em_estro:.*
       8:   em_htc:.*
       9:   gb_est1:.*
      10:   gb_est2:.*
      11:   gb_htc:.*
      12:   gb_est3:.*
      13:   gb_est4:.*
      14:   gb_est5:.*
      15:   em_estom:.*
      16:   em_estfun:.*
      17:   gb_gss:.*
      18:   em_gss_hum:.*
      19:   em_gss_inv:.*
      20:   em_gss_pln:.*
      21:   em_gss_vrt:.*
      22:   em_gss_fun:.*
      23:   em_gss_mam:.*
      24:   em_gss_mus:.*
      25:   em_gss_other:.*
      26:   em_gss_pro:.*
      27:   em_gss_rod:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1 c	55	52.9	636	17	BH603995	BH603995 BOGQI93TF
2	55	52.9	774	13	BI157225	BI157225 60292342
3	54	51.9	691	13	BJ069595	BJ069595 BJ069595
4 c	52.5	50.5	948	17	CNS06YWS	AL421586 T3 end of
5 c	51	49.0	230	9	AJ504767	AJ504767 AJ504767
6 c	51	49.0	463	14	BQ622211	BQ622211 fchic.pk0
7	51	49.0	494	9	AI236274	AI236274 EST232836
8 c	51	49.0	499	13	BI714045	BI714045 ie31e06.x
9	51	49.0	544	17	AZ720285	AZ720285 RPCI-24-9
10 c	51	49.0	564	17	BH570656	BH570656 BOHMO30TR
11	51	49.0	730	17	AG019734	AG019734 Homo sapi
12 c	50	48.1	226	17	BH646762	BH646762 BOMFC55TF
13	50	48.1	341	9	AV207860	AV207860 AV207860
14	50	48.1	364	9	AI762240	AI762240 wh63b10.x
15	50	48.1	426	9	AA159360	AA159360 zo78f05.s
16	50	48.1	518	17	BH548953	BH548953 BOGMD10TR
17 c	50	48.1	577	17	BH708571	BH708571 BOHUQ70TR
18 c	50	48.1	584	17	AQ156697	AQ156697 nbxb0008E
19 c	50	48.1	603	17	BH492663	BH492663 BOGEI14TF
20	50	48.1	641	17	BH429828	BH429828 BOGEV19TF
21 c	50	48.1	647	17	AZ639851	AZ639851 1M0501L06
22 c	50	48.1	675	17	BH418560	BH418560 BOGPD46TF
23	50	48.1	678	17	AG084609	AG084609 Pan trogl
24 c	50	48.1	710	17	BH438591	BH438591 BOHEK07TF
25 c	50	48.1	715	17	BH578695	BH578695 BOHCX79TF
26	50	48.1	760	17	BH553018	BH553018 BOHIR13TF
27	50	48.1	775	17	BH521366	BH521366 BOHQT04TF
28	50	48.1	790	17	BH429213	BH429213 BOHCE68TR
29 c	50	48.1	800	17	BH696657	BH696657 BOMIL10TF
30	50	48.1	817	17	BH605581	BH605581 BOGPS47TR
31	50	48.1	818	17	BH461814	BH461814 BOGPN88TF
32	50	48.1	832	17	BH445663	BH445663 BOHKG72TF
33	50	48.1	835	17	BH587383	BH587383 BOGHB36TR
34	50	48.1	836	17	BH555476	BH555476 BOGYJ54TF
35 c	50	48.1	865	13	BI833209	BI833209 603087266
36	50	48.1	939	14	BQ890549	BQ890549 AGENCOURT
37 c	50	48.1	973	13	BM450250	BM450250 AGENCOURT
38	49.5	47.6	591	13	BM230202	BM230202 K0291G03-
39	49.5	47.6	764	12	BG173055	BG173055 60235539
40 c	49	47.1	235	10	AV979344	AV979344 AV979344
41	49	47.1	313	17	B23342	B23342 F28G6TR IGF
42	49	47.1	351	10	AW296496	AW296496 UI-H-BW0-
43 c	49	47.1	360	17	AZ467844	AZ467844 1M0279G01
44	49	47.1	417	10	BB778886	BB778886 BB778886
45	49	47.1	423	10	AV750133	AV750133 AV750133

ALIGNMENTS

RESULT 1
BH603995/c
LOCUS   BH603995   636 bp   DNA   linear   GSS 15-DEC-2001
DEFINITION   BOGQI93TF BOGQ Brassica oleracea genomic clone BOGQI93, DNA
             sequence.
ACCESSION   BH603995
VERSION   BH603995.1 GI:17856441
KEYWORDS   GSS.
SOURCE   Brassica oleracea.
ORGANISM   Brassica oleracea
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 636)
AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

```
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSS: BOGI93TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
source      1..636
            /organism="Brassica oleracea"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone_lib="BOGQ"
            /note="vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT    203 a    124 c    141 g    168 t
ORIGIN

Alignment Scores:
Pred. No.:                            636
Score:                  38.9          Matches:          11
Percent Similarity:     55.00         Conservative:     2
Best Local Similarity:  86.67%        Mismatches:       2
Query Match:            73.33%        Indels:           0
DB:                     52.88%        Gaps:             0
                        17

US-09-833-017B-4 (1-21) x BH603995 (1-636)

Qy    2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThr 16
         ||||||||::::||| |||::||| |||||||||||| ||||||||
Db  589 GGTAGTTTAAATACCCATTTCAACTATTAACCCTTCTTTCACA 545

RESULT 2
BI157225                  774 bp    mRNA    linear   EST 05-JUL-2001
LOCUS       BI157225
DEFINITION  602923432F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5063670 5',
            mRNA sequence.
ACCESSION   BI157225
VERSION     BI157225.1  GI:14617226
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 774)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1171 row: p column: 07
            High quality sequence stop: 769.
FEATURES    Location/Qualifiers
source      1..774
            /organism="Mus musculus"
            /strain="129,C57BL/6J,FVB/N"
            /db_xref="taxon:10090"
            /clone="IMAGE:5063670"
            /clone_lib="NCI_CGAP_Mam3"
            /tissue_type="tumor, gross tissue"
            /dev_stage="10 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
            Library constructed by Lothar Hennighausen/Chu-Xia Deng, NIH
            providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
            Reference for transgenic model: Xu et al., Nature Genetics
            22, 37-43 (1999)."
BASE COUNT    146 a    267 c    236 g    124 t
ORIGIN

Alignment Scores:
Pred. No.:                            774
Score:                  48.8          Matches:          11
Percent Similarity:     55.00         Conservative:     2
Best Local Similarity:  65.00%        Mismatches:       7
Query Match:            55.00%        Indels:           0
DB:                     52.88%        Gaps:             0
                        13

US-09-833-017B-4 (1-21) x BI157225 (1-774)

Qy    2 GlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
         |||||||  ||||||| :::|| ||||||||| ||| |||          |||:::
Db  666 GGCAGCTGCGTGACTTTTTACCGGCCCTTCAACCGCAAGTTCTACCAGATGACCGGCCGC 725

RESULT 3
BJ069595
LOCUS       BJ069595              691 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ069595 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL053010 5', mRNA sequence.
ACCESSION   BJ069595
VERSION     BJ069595.1  GI:17497955
KEYWORDS    EST.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 691)
AUTHORS     Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
TITLE       Expressed genes in X. laevis embryo
JOURNAL     Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES    Location/Qualifiers
source      1..691
            /organism="Xenopus laevis"
            /db_xref="taxon:8355"
            /clone="XL053010"
            /clone_lib="NIBB Mochii normalized Xenopus tailbud
            library"
            /tissue_type="whole embryo"
            /dev_stage="stage 25"
            /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
            were oligo-dT primed and directionally cloned. Staging
            according to Nieuwkoop and Faber. Library is subtracted
            and was constructed by N. Garrett and A.M. Zorn,
            (Wellcome/CRC Institute)."
BASE COUNT    173 a    148 c    206 g    162 t
ORIGIN

Alignment Scores:
Pred. No.:                            691
Score:                  62.8          Matches:          10
Percent Similarity:     54.00         Conservative:     2
Best Local Similarity:  70.59%        Mismatches:       5
Query Match:            58.82%        Indels:           0
                        51.92%        Gaps:             0
```

```
DB:                              13         Gaps:                    0

US-09-833-017B-4 (1-21) x BJ069595 (1-691)

Qy       1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
            |||  |||::::      ||||||||||||||  |||||||:::
Db     493  AGCAGCTCGATTGAAGAATTCTTCCGACTCATGAATTGGCAATTCACGGAG 543

RESULT 4
CNS06YWS/c
LOCUS          CNS06YWS                948 bp    DNA        linear   GSS 06-JUL-2001
DEFINITION     T3 end of clone AY0AA013F07 of library AY0AA from strain CBS 6340
               of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION      AL421586
VERSION        AL421586.1  GI:12204785
KEYWORDS       GSS.
SOURCE         Kluyveromyces thermotolerans.
ORGANISM       Kluyveromyces thermotolerans
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE      1  (bases 1 to 948)
AUTHORS        Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
               Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
               de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
               Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
               Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
               Wincker,P. and Weissenbach,J.
TITLE          Genomic exploration of the hemiascomycetous yeasts: 1. A set of
               yeast species for molecular evolution studies
JOURNAL        FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE        20584711
PUBMED         11152876
REFERENCE      2  (bases 1 to 948)
AUTHORS        Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
               and Dujon,B.
TITLE          Genomic exploration of the hemiascomycetous yeasts: 10.
               Kluyveromyces thermotolerans
JOURNAL        FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE        20584720
PUBMED         11152885
REFERENCE      3  (bases 1 to 948)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
               2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
               seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT        This GSS is part of a random genomic sequencing program of thirteen
               yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
               exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
               Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
               lactis var. lactis, Debaryomyces hansenii var. marxianus Pichia
               angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
               Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
               5 kb were prepared and both extremities were sequenced. See
               keywords for description of this sequence and for the sequence of
               the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1..948
                     /organism="Kluyveromyces thermotolerans"
                     /strain="CBS 6340"
                     /db_xref="taxon:4916"
                     /clone="AY0AA013F07"
                     /clone_lib="AY0AA"
                     /note="end : T3"
     misc_feature    <290..>946
                     /note="similar to Saccharomyces cerevisiae ORF YMR001c [
                     CDC5 ; involved in regulation of DNA replication ]"
                     /evidence=not_experimental
BASE COUNT       321 a    199 c    196 g    230 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:  160          Length:          948
```

```
Score:                          52.50        Matches:                 13
Percent Similarity:             71.43%       Conservative:             2
Best Local Similarity:          61.90%       Mismatches:               5
Query Match:                    50.48%       Indels:                   1
DB:                             17           Gaps:                     1

US-09-833-017B-4 (1-21) x CNS06YWS (1-948)

Qy       1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
            |||||||||||||  |||||||||||  |||||||||   |||||
Db     819  TCAGGTTCTGTMAGTACT---TTTCGTCTTTTCAAGAGGTCCATTAGGGACCCGTTAGGG 763

Qy      21  Lys 21
            :::
Db     762  CAA 760

RESULT 5
AJ504767/c
LOCUS          AJ504767                230 bp    mRNA       linear   EST 07-AUG-2002
DEFINITION     AJ504767 Pyrus communis fruit mesocarp Pyrus communis cDNA clone
               L1041, mRNA sequence.
ACCESSION      AJ504767
VERSION        AJ504767.1  GI:22134489
KEYWORDS       EST.
SOURCE         pear.
ORGANISM       Pyrus communis
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Rosidae; eurosids I; Rosales; Rosaceae; Maloideae; Pyrus.
REFERENCE      1  (bases 1 to 230)
AUTHORS        Fonseca,S., Hackler,L., Zvara,A., Ferreira,S., Balde,A., Dudits,D.,
               Pais,M.S. and Puskas,L.G.
TITLE          Global analysis of gene expression in the ripeninig pear
JOURNAL        Unpublished (2002)
COMMENT        Contact: Fonseca SCM
               Plant Biotechnology
               Inst. de Ciencia Aplicada e Tecnologia, Ed. ICAT
               Campus da Fcul, Campo Grande, 1749-016 Lisbon, Portugal.
FEATURES             Location/Qualifiers
     source          1..230
                     /organism="Pyrus communis"
                     /cultivar="Rocha"
                     /db_xref="taxon:23211"
                     /clone="L1041"
                     /clone_lib="Pyrus communis fruit mesocarp"
                     /tissue_type="fruit mesocarp"
                     /note="country=Portugal"
BASE COUNT        71 a      41 c      56 g      62 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                          56.1         Length:                  230
Percent Similarity:             51.00%       Matches:                 11
Best Local Similarity:          76.47%       Conservative:             2
Query Match:                    64.71%       Mismatches:               4
DB:                             49.04%       Indels:                   0
                                9            Gaps:                     0

US-09-833-017B-4 (1-21) x AJ504767 (1-230)

Qy       1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGln 17
            |||||||||||||  |||||||||||  ||||  :::
Db      62  TCAGGTTCTTTGATGACATTCTTTAGGTTGCCAAACGATCTTCGACAAAG 12

RESULT 6
BQ622211/c
LOCUS          BQ622211                463 bp    mRNA       linear   EST 01-JUL-2002
DEFINITION     fchlc.pk003.112 Conidiobolus cornatus ARSEF 512 Conidiobolus
               coronatus cDNA, mRNA sequence.
ACCESSION      BQ622211
VERSION        BQ622211.1  GI:21649380
KEYWORDS       EST.
SOURCE         Conidiobolus coronatus.
```

```
ORGANISM   Conidiobolus coronatus
           Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
           Ancylistaceae; Conidiobolus.
REFERENCE  1 (bases 1 to 463)
  AUTHORS  Freimoser,F.M., Screen,S., Hu,G., and St. Leger,R.J.
  TITLE    EST analysis of genes expressed by the zygomycete pathogen
           Conidiobolus coronatus during optimized secretion of proteins
  JOURNAL  Unpublished (2002)
  COMMENT  Contact: Freimoser F. M.
           Department of Entomology
           Universty of Maryland
           4112 Plant Sciences Building, College Park, MD 20742, USA
           Tel: 301 405 16 13
           Fax: 301 314 92 90
           Email: ff34@umail.umd.edu.
FEATURES              Location/Qualifiers
     source          1..463
                     /organism="Conidiobolus coronatus"
                     /strain="ARSEF 512"
                     /db_xref="taxon:34488"
                     /clone_lib="Conidiobolus cornatus ARSEF 512"
                     /note="Vector: Unizap; Conidiobolus coronatus was grown in
                     minimal medium supplemented with Manduca sexta cuticle and
                     peptone for 18 hours. A cDNA library was constructed in
                     the unidirectional Lambda vector Unizap."
BASE COUNT      157 a     94 c     72 g    134 t
ORIGIN

Alignment Scores:
Pred. No.:               125
Score:                   51.00
Percent Similarity:      72.22%
Best Local Similarity:   55.56%
Query Match:             49.04%
DB:                      14

US-09-833-017B-4 (1-21) x BQ622211 (1-463)

Qy    3  SerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
         |||:::  |||||:::|||||||||||||:::   |||||
Db  233  TCAATTTTGAGNTTGTTCATGTTATTTTCAAGATCTTTTACAGAGCCGTTGGGT 180

RESULT 7
LOCUS       AI236274            494 bp    mRNA    linear   EST 31-JAN-1999
DEFINITION  EST232836 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVDD54 3' end, mRNA sequence.
ACCESSION   AI236274
VERSION     AI236274.1   GI:3829780
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 494)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
  COMMENT   Other_ESTs: TC61188
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source          1..494
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2041604"
                     /db_xref="taxon:10118"
                     /clone="ROVDD54"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      150 a    114 c    113 g    117 t
ORIGIN

Alignment Scores:
Pred. No.:               135
Score:                   51.00
Percent Similarity:      75.00%
Best Local Similarity:   65.00%
Query Match:             49.04%
DB:                      9

US-09-833-017B-4 (1-21) x AI236274 (1-494)

Qy    1  SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
         |||||||||||||||||::::||   ||||||  |||  |||||||||
Db  315  TCCGGATCNCTCTCACATATTCCAGACTTCCAACCTATCAAAAACTGGAGCACTGGT 374

RESULT 8
LOCUS       BI714045            499 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  ie3le06.x1 Kaestner ngn3 wt Mus musculus cDNA 3' similar to
            TR:Q9Z1Q8 Q9Z1Q8 HYPOTHETICAL 94.9 KD PROTEIN. ;, mRNA sequence.
ACCESSION   BI714045
VERSION     BI714045.1   GI:15689740
KEYWORDS    EST.
SOURCE      Mus musculus
            house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
  COMMENT   Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,
            2000) Library was constructed by Catherine Lee DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Marie Scearce
            (mscearce@mail.med.upenn.edu)
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 427.
FEATURES              Location/Qualifiers
     source          1..499
                     /organism="Mus musculus"
                     /strain="129/Sv x CD1"
                     /db_xref="taxon:10090"
                     /clone_lib="Kaestner ngn3 wt"
                     /dev_stage="p.c. 14.5"
                     /lab_host="E. coli-DH12S (GIBCO)"
                     /note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site_1:
                     Not I; Site_2: Sal I; The library was prepared by
                     Catherine S. Lee and has not been published. The pancreas
                     was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,
                     2000). The cDNA's were prepared with an oligo containing a
                     NotI site, and SalI linkers were added to the ends. The
```

```
                              inserts were cut with NotI before being cloned into the
                              NotI-SalI sites in the vectors. This is one of two
                              libraries, ngn3 wt and ngn3 -/-. The wt library is in
                              pSPORT1, T7 promoter is 5'."

BASE COUNT        144 a    104 c    127 g    124 t
ORIGIN

Alignment Scores:
Pred. No.:               136                          Length:               499
Score:                   51.00                        Matches:              12
Percent Similarity:      70.59%                       Conservative:         0
Best Local Similarity:   70.59%                       Mismatches:           5
Query Match:             49.04%                       Indels:               0
DB:                                                   Gaps:                 0

US-09-833-017B-4 (1-21) x BI714045 (1-499)

Qy       5 SerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
           ||||||  ||||||||  ||||||||||  |||||||||
Db      59 AGTACAGTTCAGACTGGAAAAATAGCTCTTTCACTATTCTTTGGGAAAA 9

RESULT 9
LOCUS        AZ720285               544 bp    DNA        linear   GSS 24-JAN-2001
DEFINITION   RPCI-24-920I.TJ RPCI-24 Mus musculus genomic clone RPCI-24-920I,
             DNA sequence.
ACCESSION    AZ720285
VERSION      AZ720285.1  GI:12461826
KEYWORDS     GSS.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 544)
AUTHORS      Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
             Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
             Russell,D., de Jong,P. and Fraser,C.M.
TITLE        Mouse BAC End Sequences from Library RPCI-24
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: RPCI-24-920I.TV
             Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: szhao@tigr.org
             Clones are derived from the mouse BAC library RPCI-24. For BAC
             library availability, please contact Pieter de Jong
             (pdejong@mail.cho.org). Clones may be purchased from BACPAC
             Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
             page: http://ww.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
             Plate: 92  row: O  column: 1
             Seq primer: SP6
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..544
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-920I"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamHl; Site_2: BamHl;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamHl sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       175 a    114 c    129 g    126 t
ORIGIN

Alignment Scores:
```

```
Pred. No.:               151                          Length:               544
Score:                   51.00                        Matches:              9
Percent Similarity:      80.00%                       Conservative:         3
Best Local Similarity:   60.00%                       Mismatches:           3
Query Match:             49.04%                       Indels:               0
DB:                                                   Gaps:                 0

US-09-833-017B-4 (1-21) x AZ720285 (1-544)

Qy       6 ThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
           |||  ||||||||  ||::::::|||  |||::::|||||
Db      19 ACAATATTCCGCTTGTTAAATAAGACTTCAATCAGAGTCTAGGC 63

RESULT 10
LOCUS        BH570656               564 bp    DNA        linear   GSS 14-DEC-2001
DEFINITION   BOHMO30TR BOHM Brassica oleracea genomic clone BOHMO30, DNA
             sequence.
ACCESSION    BH570656
VERSION      BH570656.1  GI:17822495
KEYWORDS     GSS.
SOURCE       Brassica oleracea.
ORGANISM     Brassica oleracea
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE    1  (bases 1 to 564)
AUTHORS      Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE        Whole genome shotgun sequencing of Brassica oleracea
JOURNAL      Unpublished (2001)
COMMENT      Other_GSSs: BOHMO30TF
             Contact: Chris Town
             TIGR
             9712 Medical Center Drive, Rockville, MD 20850, USA.
             Tel: 301-838-3523
             Fax: 301-838-0208
             Email: cdtown@tigr.org
             DNA is from a doubled haploid provided by Tom Osborn.
             Seq primer: TR
             Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..564
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHMO30"
                     /clone_lib="BOHM"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       157 a    163 c    120 g    124 t
ORIGIN

Alignment Scores:
Pred. No.:               157                          Length:               564
Score:                   51.00                        Matches:              11
Percent Similarity:      73.68%                       Conservative:         3
Best Local Similarity:   57.89%                       Mismatches:           5
Query Match:             49.04%                       Indels:               0
DB:                                                   Gaps:                 0

US-09-833-017B-4 (1-21) x BH570656 (1-564)

Qy       1 SerGlySerLeuSerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeu 19
           ||||||  ||||||  |||||| ||||||  ||::::||
Db     323 TCAGGTGAGCTTTCTCCTTCTCCTTGTTCTCCTGTCTCCTCCATTCTCCTCTCTCTC 267

RESULT 11
LOCUS        AG019734               730 bp    DNA        linear   GSS 16-OCT-1999
DEFINITION   Homo sapiens genomic DNA, 21q region, clone: B2289H10 N056(RP),
             genomic survey sequence.
ACCESSION    AG019734
VERSION      AG019734.1  GI:6045678
```

```
KEYWORDS    GSS.
SOURCE      Homo sapiens DNA, clone:B2289H10 N056(RP).
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 730).
AUTHORS     Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
            Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens genomic DNA, chromosome 21q
JOURNAL     Published Only in DataBase (1999)
REFERENCE   2 (bases 1 to 730)
AUTHORS     Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
            Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-OCT-1999) Masahira Hattori, RIKEN Genomic Sciences
            Center(GSC) c/o Kitasato University; 1-15-1 Kitasato, Sagamihara
            228-8555, Japan (E-mail:hattori@hgc.ims.u-tokyo.ac.jp,
            Tel:042-778-9923, Fax:042-778-9924)
FEATURES             Location/Qualifiers
     source          1..730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q"
                     /clone="B2289H10 N056(RP)"
BASE COUNT     192 a    141 c    127 g    270 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                    211
Percent Similarity:     51.00%
Best Local Similarity:  70.00%
Query Match:            55.00%
DB:                     49.04%
                           17

Length:               730
Matches:               11
Conservative:           3
Mismatches:             6
Indels:                 0
Gaps:                   0

US-09-833-017B-4 (1-21) x AG019734 (1-730)

Qy    1 SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly 20
          ::|||||||::|||||| |||||||| |||::: |||  |||||
Db  370 ACAGGATCCATTCCACAATGAATTCCTTATTCAATATTTCTTGGGTCCAAGGATTAGGA 429

RESULT 12
LOCUS       BH646762
BH646762/c
DEFINITION  BOMFC55TF BO_2_3_KB Brassica oleracea genomic clone BOMFC55, DNA
            sequence.                                              226 bp    DNA    linear    GSS 19-FEB-2002
ACCESSION   BH646762
VERSION     BH646762.1  GI:18704175
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 226)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..226
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMFC55"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      47 a     50 c     54 g     75 t
ORIGIN

Alignment Scores:
Pred. No.:                226
Score:                    80.7
Percent Similarity:     50.00
Best Local Similarity:  76.47%
Query Match:            58.82%
DB:                     48.08%
                           17

Length:               226
Matches:               10
Conservative:           3
Mismatches:             4
Indels:                 0
Gaps:                   0

US-09-833-017B-4 (1-21) x BH646762 (1-226)

Qy    5 SerThrPhePheArgLeuPheAsnArgSerPheThrGlnAlaLeuGlyLys 21
          :::|||:::: ||||: |||||| |||||::||| ||||||||
Db   64 GCAACTTACCAACGACTCGTGAACCGTATGTTCTCAACAACTCGGTAAA 14

RESULT 13
AV207860
LOCUS       AV207860                341 bp    mRNA    linear    EST 30-OCT-1999
DEFINITION  AV207860 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700099M17 3', mRNA sequence.
ACCESSION   AV207860
VERSION     AV207860.1  GI:6148713
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 341)
AUTHORS     Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,M., Tsunoda,T.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Konno,H., et al 1999)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..341
                     /organism="Mus musculus"
```

```
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="1700099M17"
                /clone_lib="RIKEN full-length enriched, adult male testis"
                /sex="male"
                /tissue_type="testis"
                /dev_stage="adult"
                /lab_host="SOLR"
                /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                prepared and sequenced in Mouse Genome Encyclopedia
                Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in
                RIKEN. Division of Experimental Animal Research in Riken
                contributed to prepare mouse tissues. 1st strand cDNA was
                primed with a primer [5'
                GAGAGAGAAGAAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                prepared by using trehalose thermo-activated reverse
                transcriptase and subsequently enriched for full-length by
                cap-trapper. Second strand cDNA was prepared with the
                primer adapter of sequence [5'
                GAGAGAGAGAGCGGCCGCCAATTAATTCTCGAGTTAATTAATTAATTAATCCCCCCCCCC
                3']. cDNA was cloned into the XhoI and BamHI sites."
BASE COUNT     127 a      66 c      41 g     107 t
ORIGIN

Alignment Scores:
Pred. NO.:            129             Length:             341
Score:               50.00           Matches:             10
Percent Similarity:  70.59%          Conservative:        2
Best Local Similarity: 58.82%        Mismatches:          5
Query Match:         48.08%          Indels:              0
DB:                                  Gaps:                0

US-09-833-017B-4 (1-21) x AV207860 (1-341)

Qy     1  SerGlySerLeuSerThrPheArgLeuPheAsnArgSerPheThrGln  17
          ||||||  ||  |  ||||||||||||  |||||   ::|||:|||||:
Db   184  AGTGGAAGAAGAAGTAAATTTTTTAGATACTTTAATGCAGCTTTTTCTCAA  234

RESULT 14
AI762240
LOCUS       AI762240              364 bp    mRNA    linear    EST 20-DEC-1999
DEFINITION  wh63b10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2385403 3',
            mRNA sequence.
ACCESSION   AI762240
VERSION     AI762240.1  GI:5177907
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 364)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1075     Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2385403"
```

```
                /clone_lib="NCI_CGAP_Kid11"
                /lab_host="DH10B"
                /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                prepared, and ss circles were made in vitro. Following HAP
                purification, this DNA was used as tracer in a subtractive
                hybridization reaction. The driver was PCR-amplified cDNAs
                from a pool of 5,000 clones made from the same library
                (cloneIDs 1322376-1323911, 1456007-1456775, and
                1500552-1502855). Subtraction by Bento Soares and M.
                Fatima Bonaldo."
BASE COUNT     41 a      98 c      47 g     178 t
ORIGIN

Alignment Scores:
Pred. NO.:           139             Length:             364
Score:               50.00           Matches:             10
Percent Similarity:  66.67%          Conservative:        2
Best Local Similarity: 55.56%        Mismatches:          6
Query Match:         48.08%          Indels:              0
DB:                                  Gaps:                0

US-09-833-017B-4 (1-21) x AI762240 (1-364)

Qy     3  SerLeuSerThrPheArgLeuPheAsnArgSerPheThrGlnAlaLeuGly  20
          |||||||  ||||||||||||  ||  |||||:  |||:   |||
Db   122  TCTCTCCTGTCTTTTTTCAGGCTTTTACAACTACTTTTTCAGTGACCTTTCCGC  175

RESULT 15
AA159360
LOCUS       AA159360              426 bp    mRNA    linear    EST 08-NOV-1997
DEFINITION  zo78f05.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:593025 3', mRNA sequence.
ACCESSION   AA159360
VERSION     AA159360.1  GI:1734171
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 426)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 768     Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 326.
FEATURES             Location/Qualifiers
     source          1..426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:593025"
                     /clone_lib="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT     120 a      89 c      59 g     157 t
```

ORIGIN

Alignment Scores:
Pred. No.:   167   Length: 426
Score: 50.00   Matches: 11
Percent Similarity: 66.67%   Conservative: 3
Best Local Similarity: 52.38%   Mismatches: 7
Query Match: 48.08%   Indels: 0
DB: 9   Gaps: 0

US-09-833-017B-4 (1-21) x AA159360 (1-426)

Qy   1 SerGlySerLeuSerThrPheArgPheAsnArgSerPheThrGlnAlaLeuGly 20
           || ::||||||:::|||||||||||| ::: ||||||
Db   23 TCAGTAGCATTATCCAGCTTTTTCCGTTTGTTCCAGAAAATAGTTCTTACTGGCCTTGGT 82

Qy   21 Lys 21
           |||
Db   83 AAA 85

Search completed: November 8, 2002, 23:39:49
Job time : 830.582 secs